ROCKWOOL FOUNDATION
BERLIN
INSTITUTE FOR THE ECONOMY AND THE FUTURE OF WORK

CReAM | Centre for Research & Analysis of Migration

DISCUSSION PAPER SERIES 18/24

Labor Market Power, Self-Employment, and Development

Francesco Amodio, Pamela Medina, Monica Morlacco

SEPTEMBER 2024

Labor Market Power, Self-Employment, and Development*

Francesco Amodio[†] Pamela Medina[‡] Monica Morlacco[§]

September 9, 2024

Abstract

This paper shows that self-employment shapes labor market power in low-income countries, affecting industrial development. Using Peruvian data, we show that wage-setting power increases with concentration, but less so where self-employment is more prevalent. A general equilibrium model shows that while concentration increases oligopsony power, it also raises labor supply elasticity by pushing workers into self-employment, thereby mitigating labor market power. Conversely, pro-competitive policies that draw workers into salaried jobs may increase labor market power, with limited overall impact. We demonstrate that these policies are only effective if they tackle labor market power.

Keywords: labor market power, monopsony, self-employment, sorting, development.

JEL Codes: J2, J3, J42, L10, O14, O54.

*We thank Jacob Schmidt and Paola Marino Moreno for excellent research assistance. For their insightful comments and suggestions, we thank Vittorio Bassi, Robin Burgess, Paula Bustos, Stefano Caria, Arin Dube, Rafael Dix-Carneiro, Marcela Eslava, Mayara Felix, Fernando Fernandez, Nicolas Gendron-Carrier, Kyle Herkenhoff, Jonas Hjort, Rocco Macchiavello, Isabela Manelici, Marti Mestieri, Suresh Naidu, Tommaso Porzio, Todd Schoellman, Gabriel Ulyssea, Jose Vasquez, Daniel Xu, and all participants to seminars at Bologna, BU, CEMFI, CREST, Dartmouth, Duke, EIEF, JILAEE-UChicago, Loughborough, LSE, Manchester, McGill, McMaster, Milan, Minneapolis Fed, Montevideo, NCID-Navarra, Nottingham, Notre Dame, SFU, Stockholm, UAB, ULosAndes, UBC, UChile, UC Riverside, UC San Diego, UofT, UPF, USC, Warwick, Western, and many other conferences and workshops. Errors remain our own. We gratefully acknowledge financial support from the CEPR/DFID PEDL initiative (ERG 6500), McGill University and the Social Sciences and Humanities Research Council (SSHRC) of Canada - Insight Grant Program.

[†] francesco.amodio@mcgill.ca, McGill University and IZA, Department of Economics and ISID, 855 Sherbrooke St. West, Montreal, QC, H3A 2T7.

[‡] pamela.medinaquispe@rotman.utoronto.ca, University of Toronto, UTSC and Rotman School of Management, 105 St George St, Toronto, ON M5S 3E6.

[§] morlacco@usc.edu, University of Southern California, Department of Economics, 3620 S. Vermont Avenue, Los Angeles, CA 90089-1422.

1 Introduction

Millions of people in low- and middle-income countries rely on subsistence labor for their livelihoods. Still, the role of informal self-employment in economic development remains a contentious issue. The traditional view is that economic development stems from the modern industrial sector, and that informal self-employment is bound to disappear as formal manufacturing expands (Lewis, 1954; Harris and Todaro, 1970; Rauch, 1991). As a result, the creation of manufacturing salaried jobs has become a cornerstone of industrial development policy (UN General Assembly, 2015). Despite these efforts, self-employment remains high in emerging countries (La Porta and Shleifer, 2014; Gollin, 2008; Poschke, 2022) and employment at larger firms stagnates (Hsieh and Olken, 2014; McMillan and Zeufack, 2022), even as GDP per capita increases. Understanding why self-employment persists and why manufacturing firms cannot absorb more workers is crucial in determining the future development trajectory of these countries and the scope for policy intervention.

Labor market structure is a potentially important but often overlooked factor influencing these outcomes. Multiple barriers to firm growth, such as high entry costs, a shortage of skilled labor, and inadequate infrastructure, result in the concentration of employment among a small number of firms (Djankov et al., 2002; Rud and Trapeznikova, 2021; Hjort, Malmberg and Schoellman, 2022). These firms may internalize their impact on local labor market conditions, reducing job opportunities and wages to increase profits. However, self-employment represents a valuable outside option for workers. Within a local labor market, workers can easily switch between self-employment and wage work (Donovan, Lu and Schoellman, 2023), and they can opt for self-employment when posted wages are too low (Blattman and Dercon, 2018; Breza, Kaur and Shamdasani, 2021).

This paper argues that understanding the interplay between labor market power and self-employment is crucial to explain the persistently high rates of self-employment in emerging economies and why development policies aimed at boosting industrial wage employment often fall short of their objectives.[1] To support these claims, we provide new evidence from Peru, an original theoretical framework, and counterfactual policy experiments. Peru serves as a meaningful case study due to its high levels of employer concentration, self-employment, and worker mobility, characteristics common to many other low- and middle-income countries.

We begin by showing that labor market power is substantial in Peru. Its extent varies across local labor markets depending on employer concentration and self-employment opportunities, with the latter acting as a constraint on employers' wage-setting power.[2] We measure labor market power as the inverse elasticity of the labor supply curve faced by individual firms, a direct measure of their ability to set wages (Manning, 2003). For estimation, we use an instrumental variable strategy that constructs firm-level labor demand shifters from the staggered

[1] See McKenzie (2017) and Bandiera et al. (2022) for a review.

[2] We define a local labor market as the combination of a 2-digit industry and a commuting zone. See Section 2 for further details and motivation for this definition.

implementation of a rural electrification program across provinces and its differential impact on firms with high vs. low ex-ante constraints in accessing electricity.

We find that the average firm-level inverse labor supply elasticity across local labor markets is positive and significant, indicating substantial wage-setting power. The implied average wage markdown is 1.42, meaning that manufacturing workers receive about 70 cents as a wage for every additional dollar they produce. The markdown increases with market concentration, suggesting oligopsony power among employers. However, this positive relationship weakens in markets where self-employment is more prevalent. We find the highest markdowns in markets with high concentration and low self-employment rates, where workers receive only 57 cents for the marginal dollar they produce. Conversely, the labor supply is relatively more elastic in markets with a large self-employment sector.

An important consideration is that the *reduced-form* inverse elasticity obtained through our identification strategy does not directly correspond to the *structural* inverse elasticity, which measures the elasticity of firm-level wages to employment changes while holding competitors' wages and employment constant (Berger, Herkenhoff and Mongey, 2022). This raises the question of whether the observed heterogeneity in the inverse elasticity of labor supply across markets reflects structural features of the economy or equilibrium effects. Answering this question is crucial for policymaking, and requires a theoretical model.

We develop a general equilibrium model of Peruvian manufacturing labor markets, where employer concentration, self-employment rates, and labor market power are jointly determined. The model's first key feature is *oligopsony*. Each local labor market features a finite number of heterogeneous firms that internalize their impact on market-level labor demand and wages and make strategic decisions accordingly. The model's second key feature is a Roy's (1951) structure of *self-selection of heterogeneous workers* across wage work and self-employment based on earnings. In addition to this, the model also features oligopsony power in the product market and endogenous entry.

The theory sheds light on the structural determinants of labor market power. In equilibrium, the (payroll-weighted) average wage markdown in a local labor market is an exact function of two endogenous variables. The first one is the payroll Herfindahl-Hirschman Index (HHI), a measure of employer concentration. Concentration positively affects the markdown and captures the *demand-side* determinants of labor market power, specifically the employers' oligopsony power. The second variable is the aggregate supply elasticity of wage work, which has a negative effect. This term reflects *supply-side* forces, notably how wage changes affect the sorting of workers across wage work and self-employment. As the relative unit wage falls, more workers choose self-employment. Falling wages make it easier to push workers out of wage employment, resulting in an increase in the overall supply elasticity of wage work and a decline in the average markdown. Similar forces can also increase labor market power when wage employment becomes more remunerative.

Our framework captures the dual role played by self-employment in the presence of labor market power. It can shield workers from the wage-setting power of firms by providing a livelihood when wage opportunities are scarce. However, it can also increase labor market power when wage employment becomes more attractive, making it difficult for industrial policies to boost wage employment and wages and potentially hindering the growth prospect of countries. Through counterfactual analysis, we show that the variable elasticity channel is quantitatively essential to understand the limited impact of industrial policies in emerging economies.

We use the model to decompose the response of sectoral average earnings to economic shocks. The average wage response reflects two components: a direct effect of the shock on the efficiency unit wage and a compositional effect on the average worker ability. The change in unit wage can be further decomposed into the change in workers' marginal revenue product (MRPL) and the change in markdown. Specifically, the first effect captures how the shock changes aggregate productivity, prices, and markups.[3]

We refer to the compositional effect on workers' ability as the *selection* channel. This effect reflects the difference in efficiency between sector-switchers and sector-stayers. The strength and direction of this channel depends on the schedules of workers' comparative and absolute advantage in the two sectors and their correlation (Adão, 2016; Amodio, Alvarez-Cuadrado and Poschke, 2020). We estimate that the workers' abilities in the two sectors are highly positively correlated and more dispersed in the self-employment sector. For estimation, we impose that ability endowments are jointly log-normally distributed, allowing us to identify the relevant ability parameters from cross-sectional data on earnings and employment shares (Heckman and Sedlacek, 1985). Our estimates imply positive selection in self-employment and no selection in wage work. They also imply that the average worker has a comparative advantage in self-employment. These findings align with experimental evidence suggesting that workers prefer self-employment to industrial jobs in poor countries (Blattman and Dercon, 2018).

Given the ability distribution parameter estimates, we rely on a Method of Simulated Moments (MSM) strategy to estimate the remaining model's parameters. We discipline the model by matching moments of the cross-sectional distributions of concentration, employment shares, and earnings. We validate the model by showing that it replicates the reduced-form patterns of labor market power across local labor markets, which were not targeted for estimation.

Armed with the estimated model, we perform two counterfactual experiments. First, we evaluate the effect of labor market power on labor market outcomes in Peru by comparing our baseline economy with one where employers act as wage-takers. In the absence of labor market power, the average share of wage employment across markets is 11 percentage points higher, up to 77% from a baseline of 66%. Furthermore, average wages are 31% higher, and earnings from self-employment 27% higher, thereby widening the earnings gap between self-employed individuals and wage workers.These effects materialize through selection and changes in labor revenue productivity, which includes the general equilibrium effect on prices and markups.

[3]In the competitive self-employment sector, the markdown is always constant and equal to one, and the MRPL is only affected by output prices.

Worker self-selection stands out as a crucial margin through which labor market power decreases worker earnings in the self-employment sector.

The second objective is to investigate the significance of labor market power for industrial development policies. We examine three categories of policies aimed at expanding wage employment: (i) enhancing firm productivity through market integration or infrastructure improvement policies (Volpe Martincus, Carballo and Cusolito, 2017; Fiorini, Sanfilippo and Sundaram, 2021); (ii) reducing fixed entry costs for employers by simplifying business registration regulations (Kaplan, Piedra and Seira, 2011; Bruhn, 2011); (iii) improving workers' skills through off and on-the-job training programs (McKenzie, 2017; Alfonsi et al., 2020). We use the our model to estimate the impact of these policies on labor market outcomes. To inform the size of the policy shocks, we analyze actual policies implemented in Peru and Mexico and their reduced-form estimated effects.

We find that policy impact varies significantly across markets, with this variation being almost entirely explained by changes in labor market power and its determinants. Pro-competitive industrial policies create salaried jobs, but also make self-employment less attractive, thereby reducing the supply elasticity of wage work and possibly having anti-competitive effects in the labor market. Markdown changes explain up to 99% and 88% of the across-market variation in the policy's impact on wages and wage employment share, respectively—making it clear that a policy is only effective if it tackles markdowns. These insights are crucial for policymakers aiming to design impactful interventions for industrialization and inclusive economic growth.

Our analysis focuses on the manufacturing sector and the role of labor market structure in shaping industrial development. Our motivation lies in the significant historical role that manufacturing has played in development. Of all the countries in the world that have managed to escape poverty, the vast majority achieved this by becoming highly industrialized, with manufacturing absorbing large amounts of unskilled labor into high-productivity work. Most of the poor countries in the world, however, are on a different trend. Despite an increase in manufacturing value-added, employment at large firms fails to expand, and self-employment rates within the manufacturing sector remain persistently high (Alfaro et al., 2023; Huneeus and Rogerson, 2023). This divergence is notable, especially considering that industrial policy in these countries primarily targets manufacturing (Juhász, Lane and Rodrik, 2023). We consider this a critical issue and a compelling reason to focus on labor market power in the manufacturing sector and its implications for industrial development and policy.

Finally, our framework provides a new lens to understand and explain the earning gap between self-employed and wage workers. Self-employed workers account for about half of the workforce in developing countries, but earnings are typically higher in wage employment (Fields, 2012). Our theory nests a model of worker sorting in a general equilibrium framework where the earning gap arises from differences in both unit earnings and selection patterns across sectors. We quantify the role of labor market power and these separate channels for the size of the gap and how it changes with policy. Yet, our framework does not contemplate other features of self-employment and wage work that could make them more or less attractive for workers such as the flexibility of self-employment, employer-mandated health insurance, and working conditions in general. Moreover, a significant share of self-employment occurs in services. Addressing these issues is a logical next step for this research.

Related Literature This paper contributes to several strands of the literature. First, we contribute to the literature on informal self-employment in low-income countries. The traditional "dual" view suggests that medium and large formal firms and informal micro-enterprises are fundamentally different and operate in entirely different economic spheres.[4] Our study challenges this view, building on the work of Maloney (1999), Ulyssea (2018), and Donovan, Lu and Schoellman (2023), among others, who show that formal and informal firms coexist in the same local labor markets, with frequent worker transitions between the two sectors. We emphasize the role of worker sorting for labor market power and outcomes in emerging countries, as well as the persistently high prevalence of self-employment in these contexts.

Second, our paper contributes to the growing literature on labor market power. Recent evidence shows that U.S. employers enjoy some degree of market power in the labor market. Several studies use employer concentration as a proxy for labor market power showing that it correlates negatively with wages (Azar, Marinescu and Steinbaum, 2022; Benmelech, Bergman and Kim, 2022). Yet, using matched employer-employee data from Oregon, Bassier, Dube and Naidu (2022) find no evidence that labor supply elasticities decrease with concentration. Similarly, in U.S. manufacturing, Yeh, Macaluso and Hershbein (2022) find that wage markdowns and employer concentration moved along different trends over the last decades. We introduce an original micro-foundation for the firm-level labor supply curve based on the self-selection of heterogeneous workers between wage work and self-employment.[5] We therefore consider both demand- and supply-side determinants of labor market power to show that, with sorting, employer concentration has a non-linear relationship with labor market power, providing a rationale for the mixed findings in the literature.

The literature on labor market power in lower-income countries is more limited. Amodio and De Roux (2022) use plant and customs data from Colombia to estimate firms' wage-setting power, concluding that workers produce 40% more than their wage level. Felix (2022) studies the impact of trade liberalization on concentration and wages in Brazil, estimating high levels of labor market power before the 1990s liberalization, but minor labor market power effects of trade.[6] She also finds that firms in local labor markets where self-employment is more prevalent face more elastic labor supply curves. In Costa Rica, Alfaro-Ureña, Manelici and Vasquez

[4]Early contributors to this literature include Lewis (1954), Harris and Todaro (1970), and Rauch (1991). See also La Porta and Shleifer (2014) for a review article.

[5]Kahn and Tracy (2024) study how local monopsony power affects the cross-sectional spatial distribution of wages and rents across cities incorporating as an extension worker sorting across sectors *à la* Roy.

[6]See also Pham (2023), MacKenzie (2021) and Gutiérrez (2023) on the interactions between trade and labor market distortions in China, India and Australia, respectively.

(2021) find minor wage effects following multinational companies' expansion, indicating low labor market power.[7] We add to this literature by presenting new evidence on the interplay between labor market power, concentration, and self-employment. We propose and estimate a novel general equilibrium model to demonstrate that self-employment acts as a check on employers' market power while, at the same time, undermining development policies in low-income countries.

Finally, our work speaks to the extensive literature on informality in low-income countries (Ulyssea, 2020). Both Dix-Carneiro and Kovak (2019) and Ponczek and Ulyssea (2021) argue that informality acts as an "unemployment buffer" by reducing trade-induced adjustment costs in the labor market. Yet, Dix-Carneiro et al. (2024) show that unemployment buffer does not necessarily imply "welfare buffer" meaning that, in the event of a negative economic shock, welfare declines by less when informality rates are modest. Our analysis adopts the notion of informal self-employment as a potential outside option for workers, and shows it has a similar dual role in the presence of labor market power: it shields workers against the wage-setting power of employers when wages are too low, but also makes it more difficult for policies that seek to boost wage employment and wages to succeed.

The remainder of the paper is organized as follows. Section 2 introduces the data and presents the empirical facts. The model and its properties are presented in Section 3, while Section 4 discusses the model estimation procedure and results. Section 5 presents the counterfactual policy analyses. Section 6 concludes.

2 Data and Facts

The empirical analysis relies on two main datasets on firms and workers. The first dataset is the Peruvian Annual Economic Survey (*Encuesta Económica Anual*, EEA), a nationwide firm-level survey conducted annually by the national statistical agency (*Instituto Nacional de Estadística e Informática*, INEI). This dataset includes standard balance sheet information such as revenues, input expenditures, and plant locations. The survey is mandatory for firms with net sales above a certain threshold, while smaller firms are sampled. As a result, the EEA provides comprehensive coverage of medium and large firms, along with a representative sample of smaller firms. To ensure consistency across years and account for changes in the reporting threshold, we focus on manufacturing firms with net sales exceeding 2 million Peruvian Soles

[7] Still in Costa Rica, Méndez and Van Patten (2022) document the critical role of labor mobility and workers' outside option on determining the degree of monopsony power of private companies and their investment in local amenities. Outside Latin America and the Caribbean, Brooks et al. (2021) show evidence of labor market power in China and India. Muralidharan, Niehaus and Sukhtankar (2023) show through experimental evidence that the labor market effects of public employment programs in rural India are consistent with monopsony power in the private sector. In South Africa, Bassier (2023) uses a variety of worker separation designs to estimate firm-level labor supply elasticities and finds high levels of monopsony. Armangué-Jubert, Guner and Ruggieri (2024) and Amodio et al. (2024) both use World Bank Enterprise Survey data from a large set of low and middle-income countries to study the relationship between labor market power and development.

(PEN) per year—approximately 700,000 USD in 2010—over the period from 2004 to 2011. Our final dataset includes 2,473 firms and 8,138 firm-year observations.

The second data source is the Peruvian National Household Survey (*Encuesta Nacional de Hogares*, ENAHO), conducted annually by INEI. This survey is nationally and regionally representative, covering urban and rural areas across the 24 Peruvian departments and the constitutional province of Callao. It provides information on household members' socioeconomic characteristics. Individuals aged 14 and older respond to a dedicated module with questions on employment status, pay, occupation, and industry. To align with the firm-level data, we focus on the years 2004 to 2011 and restrict the sample to working-age individuals (25 to 65) who have completed their education and are not yet retired. ENAHO offers several panel versions where the same households are interviewed annually for five consecutive years; we use the 2007-2011 panel to track workers' transitions between employment states.

2.1 Definitions

We define a local labor market as a 2-digit ISIC industry within a specific geographical area. These areas are primarily defined by Peruvian province boundaries, which correspond to level 2 administrative divisions, and are subdivisions of departments. Excluding Metropolitan Lima—the province that includes the capital city—the average province has a population of approximately 114,000. Metropolitan Lima is a significant outlier, with a population of 10 million. Following Piselli (2013), we define five distinct local labor markets within Lima province. In total, we analyze data from 199 geographical units and 23 manufacturing industries.[8]

Our baseline measure of concentration is the Herfindahl-Hirschman Index (HHI) for payroll, defined as $HHI_{kt}^{wn} = \sum_{i \in k} \left(s_{ikt}^{wn}\right)^2$, where $s_{ikt}^{wn} = \frac{w_{ikt} n_{ikt}}{\sum_{i \in k} w_{ikt} n_{ikt}}$ represents firm i's share of the total payroll in local labor market k in year t. Here, w_{ikt} and n_{ikt} denote the firm's wage and employment, respectively. Index values close to one indicate that a few firms dominate a significant portion of the market payroll. We also consider the employment HHI, defined as $HHI_{kt}^{n} = \sum_{i \in k} \left(s_{ikt}^{n}\right)^2$, where $s_{ikt}^{n} = \frac{n_{ikt}}{\sum_{i \in k} n_{ikt}}$, and the number of firms in the local labor market as alternative concentration measures.[9]

In the ENAHO survey, workers are classified into four categories: own-account workers, employers, auxiliary family workers, and employees. For our analysis, we group own-account workers and employers as *self-employed workers*, while employees are categorized as *wage workers*. We exclude auxiliary family workers from our classification, as they do not report monetary compensation. Additionally, ENAHO allows us to identify informal workers. A worker is considered informal if they (i) are a wage worker without health insurance,[10] or (ii)

[8]One concern with using provinces or commuting zones as geographical units is the potential for partial labor market integration. To address this, we show that the empirical patterns in the following section hold true when using departments—each consisting of roughly nine provinces on average—as the spatial unit. These results are presented in Online Appendix Tables A.1 and A.7, as well as Figures A.4 and A.5.

[9]Online Appendix Figure A.1 demonstrates a strong correlation among these measures.

[10]Employers in Peru are legally required to provide health insurance to employees.

are self-employed but not registered with the national tax authority, do not follow required procedures, and have five or fewer employees.

2.2 Employer Concentration

Employment and wages in Peruvian local labor markets are highly concentrated among a small number of medium and large firms. Panel I of Table 1 shows that the average local labor market includes about six firms, with unweighted and payroll-weighted mean wage-bill HHIs of 0.65 and 0.37, respectively. Notably, 39% of these markets are dominated by just one medium-to-large firm, and these highly concentrated markets account for approximately 8% of the nation-wide payroll. This suggests that, despite their smaller share, these concentrated markets still significantly impact the overall payroll. Importantly, this concentration is not unique to our sample but reflects broader patterns in the economy, where medium-to-large firms similarly dominate payroll and wage employment in their respective markets.[11]

Location explains about 43% of the variation in wage-bill HHI across markets, while differences across 2-digit industries only accounts for an additional 14%.[12]

2.3 Self-Employment and Flows Into and From Wage Work

In Peruvian manufacturing, as in other low- and middle-income countries, self-employment is widespread and mainly informal (Gollin, 2008; La Porta and Shleifer, 2014). Panel II of Table 1 shows that self-employment constitutes 40% of the manufacturing workforce, wage workers account for 56%, and the remaining 4% are auxiliary family workers.[13] Over 90% of self-employment is informal, both across all industries and within manufacturing.[14] In contrast, about half of wage workers are informal, with numbers declining over time.[15]

Informality influences the prevalence of self-employment by reducing the costs of starting

[11] Online Appendix Figure A.2 shows that the wage-bill HHI distribution from our data closely aligns with the 2007 Economic Census. Online Appendix Figure A.3 and Table A.2 demonstrate strong correlations in concentration measures in both datasets, even after controlling for industry and location fixed effects.

[12] Peruvian manufacturing is not more geographically clustered than its counterparts in the UK or US. To demonstrate this, we calculated the Ellison-Glaeser index of geographic concentration using data from the 2007 Economic Census for 131 manufacturing industries at the 4-digit level (Ellison and Glaeser, 1997). The results show that about 63% of Peruvian industries have a positive EG index, indicating some degree of localization, compared to 97% in the US and 94% in the UK (Duranton and Overman, 2005).

[13] Online Appendix Table A.3 shows the employment distribution across sectors for all workers and separately for self-employed and wage workers. Self-employment generally mirrors the overall workforce, except in retail and agriculture, where it is overrepresented.

[14] Of the 40% of manufacturing workers classified as self-employed, 31% are own-account workers, and 9% are employers. Only 14% of employers—about 3% of all self-employed individuals—operate formally as registered businesses. Unregistered employers, who typically hire few workers, are excluded from the EEA. All our findings remain true when focusing solely on informal self-employment as the alternative to wage work.

[15] Overall, informal workers account for 73% of the workforce in our data, a figure close to the 80% reported by the INEI in 2007. The high rate of informal self-employment contrasts sharply with the low unemployment rate, which in our data is around 3% nationally, consistent with the 3.2-3.6% reported by the ILO for Peru during the same period (International Labour Organization, 2020).

Table 1: Summary Statistics

Variable	Mean	St. Dev.
Panel I. Manufacturing Local Labor Markets		
Number of Firms	6.39	10.37
Wage-bill HHI	0.65	0.33
Wage-bill HHI (Weighted by LLM payroll share)	0.37	0.03
Employment HHI	0.63	0.35
Employment HHI (Weighted by LLM empl. share)	0.31	0.02
Percent of LLMs with 1 firm	38.78	2.27
Payroll Share of LLMs with 1 firm	7.94	1.79
Employment Share of LLMs with 1 firm	7.80	1.23
Panel II. Manufacturing Workers		
Wage Worker	0.56	0.50
Daily Wage	31.84	31.85
Self-Employed	0.40	0.49
Daily Earnings from Self-Employment	23.06	41.31
W-S Transition	0.06	0.24
S-W Transition	0.04	0.20

Notes. This table reports summary statistics from EEA firm-level data across Peruvian local labor markets (Panel I) and from ENAHO worker-level data (Panel II), averaging across all years from 2004 to 2011. 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 ≈ 0.35 USD in 2010), and annual transitions from the wage- to self-employment sector (W-S) and vice versa (S-W).

and operating a business. These lower costs contribute to variations in self-employment rates across industries. Self-employment is more prevalent in labor-intensive industries, which constitute a substantial portion of Peru's manufacturing GDP. In these industries, physical capital is less important, credit constraints are less severe, and the potential for informality is greater. Self-employment is lower in more capital-intensive sectors such as pharmaceuticals and metals, and it is virtually non-existent in oil and petroleum manufacturing.

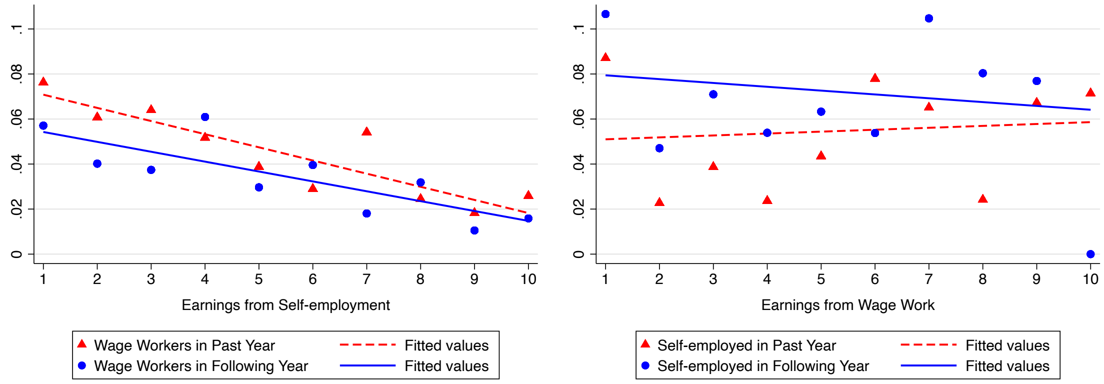
On average, earnings from self-employment are lower and more dispersed than those from wage work. Panel II of Table 1 shows that daily earnings from self-employment are approximately 28% lower than daily wages, while their standard deviation is about 30% higher.

A defining feature of labor markets in low- and middle-income countries is the high level of worker mobility between wage work and self-employment (Maloney, 1999; Donovan, Lu and Schoellman, 2023). This trend is also evident in Peru. Panel II of Table 1 shows that approximately 4% of self-employed manufacturing workers transition to wage work the following year, while 6% of manufacturing wage workers move to self-employment. When workers switch either employment status or industry, transitions between wage work and self-employment are about 20% more likely than industry changes that do not involve a status shift (56 vs. 44%).[16] Additionally, 70% of these transitions occur within the same 2-digit industry, suggesting that most moves happen within the same local labor market.[17]

[16]Although ENAHO only tracks moves without a location change, the 2007 Census shows that 85% of manufacturing workers lived in the same commuting zone as in 2002, indicating limited geographical mobility.

[17]Online Appendix Table A.3 further indicates that workers transitioning from manufacturing self-employment to wage work (or vice versa) are about four times more likely to remain in manufacturing compared to the average

Figure 1: Transition Probabilities Across the Earnings Distribution



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

Worker transitions correlate with earnings. Figure 1 shows the likelihood of switching to or from wage and self-employment across deciles of the self-employment and wage earnings distributions. The left panel indicates that workers who have recently transitioned from wage work, or are about to become wage workers, are more likely to be among the lowest-earning self-employed individuals. In contrast, the right panel of Figure 1 shows that transitions to and from self-employment are not systematically correlated with earnings from wage work.

Thus, workers at the margin between self-employment and wage work consistently earn less than inframarginal self-employed workers and have similar earnings to inframarginal wage workers. These findings seem to suggest positive selection into self-employment but no selection into wage work. We will elaborate on this point later.

2.4 Concentration, Self-Employment Rates, and Earnings

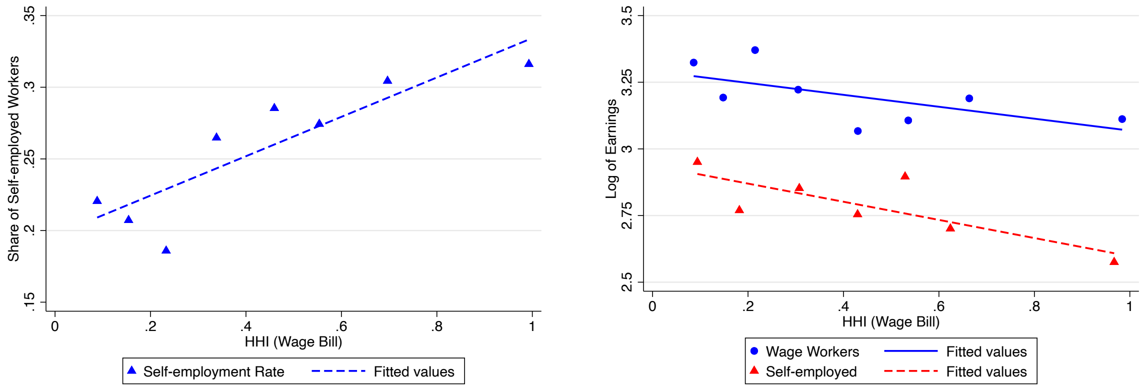
The final pattern we highlight in the data is the systematic relationship between concentration and self-employment across local labor markets. The left panel of Figure 2 shows that the average share of self-employment increases consistently across deciles of payroll HHI, indicating higher self-employment rates in more concentrated labor markets. Regression analysis further supports this correlation. We conduct a worker-level regression of a self-employment dummy on the log of wage-bill HHI in the worker's local labor market for the same year. The results, presented in Columns 1 to 3 of Online Appendix Table A.4, reveal that the relationship between concentration and self-employment is positive and significant, even after controlling for individual characteristics, industry, and location fixed effects.[18]

The right panel of Figure 2 shows the correlation between concentration and earnings from

worker switching status across all sectors.

[18]Online Appendix Tables A.5 and A.6 report coefficient estimates using employment HHI and the number of firms, respectively, as alternative concentration measures, showing similar results.

Figure 2: Concentration, Self-Employment Rate, and Earnings



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. 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.

wage work and self-employment. In markets where concentration is higher, wages tend to be lower, and self-employment is also less lucrative. The regression results, shown in Columns 4 to 9 of Online Appendix Table A.4, further support these patterns.

The decline in both the wage employment share and wages with increasing concentration is observed among both formal and informal wage workers, as shown in Online Appendix Figures A.6 and A.7. This suggests that in concentrated markets, self-employment serves as an alternative to both formal and informal wage work. The right panels in both figures also show that, despite differences in earnings levels, the wages of both formal and informal wage workers decrease at the same rate as concentration increases.

These findings reinforce the sorting narrative proposed earlier: in highly concentrated markets, where wages are lower, more workers choose self-employment. As wage workers, these individuals would have earned similar wages to their peers. However, as self-employed workers, they tend to earn less than their peers, leading to a decrease in average earnings from self-employment as market concentration increases.

2.5 Labor Market Power

The observed co-movements between concentration, self-employment, and earnings raise questions about the role of labor market power in Peruvian labor markets. While employer concentration is negatively associated with wages, this relationship alone does not prove labor market power, as both concentration and wages are equilibrium outcomes. To pin down labor market power, we estimate the inverse elasticity of labor supply faced by individual firms (Manning, 2003). By examining how this varies with labor market concentration and self-employment rates, we can gain deeper insights into the role of labor market power in this context.

Empirical Strategy We estimate the following regression model:

$$\ln w_{i(j,g)t} = \beta \ln l_{i(j,g)t} + \alpha_i + \eta_{(j,g)t} + u_{i(j,g)t}, \tag{1}$$

where $w_{i(j,g)t}$ is the wage paid by firm i in year t in its local labor market, defined by a manufacturing industry j within a province or commuting zone g, and $l_{i(j,g)t}$ is employment at the same firm. α_i is a firm fixed effect that captures differences across firms that do not change over time. $\eta_{(j,g)t}$ is a market $\times$ year fixed effect that accounts for aggregate yearly shocks at the local labor market level. This allows β to measure the firm-specific inverse labor supply elasticity of wage work while holding the aggregate labor supply constant.

To estimate the parameters in equation (1) consistently, we require a firm-level labor demand shifter, as OLS estimates may be biased due to the interdependence of wages and employment. We address this by using the rollout of the Rural Electrification Program (*Programa de Electrificación Rural*, PER), launched by the Peruvian Ministry of Energy and Mining in 1993 to foster economic and social growth in rural areas (Dasso and Fernandez, 2015). Between 1994 and 2012, the program implemented 628 projects across rural Peru, prioritizing districts with high poverty rates, low electricity coverage, and high renewable energy potential, with a total investment of USD 657.5 million (Dasso, Fernandez and Ñopo, 2015).

Our approach builds on the idea that electrification through PER increased firms' marginal productivity and labor demand, especially for firms previously facing greater constraints in accessing electricity (Abeberese, Ackah and Asuming, 2019). To operationalize this approach, we first create the variable PER_{gt}, equal to the cumulative number of completed PER projects in location g up to year t. We then follow Bau and Matray (2023) to identify firms facing electricity access constraints at baseline.

For a firm i in market (j, g) producing output $y_{i(j,g)t}$ at time t and selling it in an imperfectly competitive market, the unit price $p_{i(j,g)t}$ is a markup $\mu_{i(j,g)t}$ over marginal cost. The firm uses a Cobb-Douglas production function with industry-specific input elasticities, where θ^e_j represents the output elasticity of electricity. The shadow cost of electricity, varying across firms and industries, is denoted by $\tau^e_{i(j,g)t}$. The electricity revenue share at firm i is given by $\alpha^e_{i(j,g)t} = \frac{e_{i(j,g)t}}{p_{i(j,g)t}y_{i(j,g)t}}$, where $e_{i(j,g)t}$ is total electricity bill.

Profit maximization implies $\frac{\theta^e_j}{\alpha^e_{i(j,g)t}} = \mu_{i(j,g)t}(1 + \tau^e_{i(j,g)t})$, which we can rewrite as:

$$\ln(\alpha^e_{i(j,g)t})^{-1} = \ln(\mu_{i(j,g)t}) + \ln(1 + \tau^e_{i(j,g)t}) - \ln \theta^e_j. \tag{2}$$

This shows that we can estimate the firm-level wedge $\tau^e_{i(j,g)t}$ as the residual from a regression of the log of the inverse electricity share of revenues on industry fixed effects and firm-level markups.[19] We include 4-digit ISIC Rev. 4 code fixed effects to control for industry-specific

[19] Intuitively, in the absence of distortions, the electricity revenue share $\alpha^e_{i(j,g)t}$ should equal the output elasticity θ^e_j. However, a firm's electricity share can fall below this optimal level if the firm either has market power, which

output elasticities and use second-degree polynomials of output market shares in both the local labor market and nationwide to flexibly account for firm-level markups.[20] To mitigate the impact of outliers and address measurement error, we create a dummy variable, $EC_{i(j,g)}$, which equals one for firms with an estimated wedge $\hat{\tau}^e_{i(j,g)t}$above the median at baseline, indicating tighter constraints in accessing electricity.[21]

The interaction $PER_{gt} \times EC_{i(j,g)}$ is our instrumental variable (IV). It combines variation in program rollout across geography and over time with variation across firms within industries in access to electricity at baseline. The first-stage regression specification is

$$\ln l_{i(j,g)t} = \gamma PER_{gt} \times EC_{i(j,g)} + \phi_i + \delta_{(j,g)t} + v_{i(j,g)t}, \tag{3}$$

with ϕ_i and $\delta_{(j,g)t}$ capturing firm fixed effects and local labor market $\times$ year fixed effects, respectively, following the second-stage regression specification in equation (1).

The validity of this IV approach relies on three key assumptions. First, the instrument must be strongly correlated with employment, which holds if the electrification program boosts labor demand, particularly for firms with limited electricity access. Second, the instrument must be orthogonal to the wage and employment trends of electricity-constrained firms within each local labor market. This is plausible since the Ministry did not consider local firms or industries when implementing the program. Finally, the instrument must satisfy the exclusion restriction, meaning electrification should not differentially affect labor supply to electricity-constrained firms. This ensures that changes in employment and wages reflect movements along the labor supply curve, allowing us to trace out its slope. To support this assumption, we include local labor market $\times$ year fixed effects in all specifications. These fixed effects capture and control for changes in labor supply common to all firms within a market, even if these vary locally across industries. Importantly, we demonstrate below that our estimates remain robust when accounting for differences across firms over time at a more granular geographical level.

The exclusion restriction also requires that the labor demand shock does not affect wages via other channels, such as rents captured by workers.[22] While this could be a concern, it is unlikely in Peruvian manufacturing, where workers have minimal bargaining power. Union density was consistently low during the analysis period, ranging from 1.9% to 3.2%, placing Peru in the bottom 5% nations in unionization rates (International Labour Organization, 2020).

Another concern is whether the electrification program created sufficient variation across Peruvian local labor markets. To explore this, we examine districts as the geographical unit.

allows it to set a higher markup $\mu_{i(j,g)t}$, or if it faces a higher shadow cost of electricity, captured by $\tau^e_{i(j,g)t}$.

[20]Output markups can be expressed as an increasing function of a firm's output market share in many macroeconomic models, such as Atkeson and Burstein (2008). This approach also aligns with our theoretical model in Section 3. The results are robust to (i) not controlling for output market shares (implicitly assuming no market power), (ii) controlling only for local labor market shares, and (iii) controlling only for national shares.

[21]Online Appendix Figure A.9 shows the distribution of these wedges and the median value used as a cutoff.

[22]The firm's first-order condition in this case links wage markdowns, firm rents, worker bargaining power, and labor supply elasticity. See Wong (2023) for details.

Peru has 1,838 districts, each a subdivision of a province, with an average of about 9 districts per province. We find that 15% of districts in the firm-level IV estimation sample were affected by the program. These districts account for 41% of firm-level observations and 17% of the manufacturing workforce (based on ENAHO data). Online Appendix Figure A.8 offers additional details on the program's implementation. Initially, the targeted districts had a relatively low share of manufacturing employment. By the end of the period, however, the program had reached districts with a higher proportion of manufacturing employment, possibly due to the program itself spurring growth in these areas.

Results Table 2 presents the inverse elasticity IV estimates and standard errors. We report for each estimate the *F-statistic* associated with the Sanderson-Windmeijer multivariate test of excluded instruments, confirming that the instrument provides meaningful identifying variation throughout. Online Appendix Table A.8 reports the first-stage regression results.

Column 1 reports the results for the total sample of manufacturing firms. The firm-level inverse labor supply elasticity is estimated at 0.42, corresponding to a labor supply elasticity of 2.36, which is statistically significant at the 1% level. This elasticity implies a markdown of 1.42 between the marginal revenue product of labor and the wage paid. In other words, workers generate 42% more as value than what they earn at the margin, taking home 70 cents for every marginal dollar they produce.[23]

Columns 2 to 4 focus on different subsamples.[24] In Column 2, we estimate labor market power separately for markets with varying levels of labor market concentration.[25] For firms in the least concentrated labor markets ($HHI \leq 0.18$), we estimate an inverse labor supply elasticity that is statistically and economically insignificant. As concentration increases, labor market power rises. In moderately concentrated markets ($0.18 < HHI \leq 0.25$), workers take home nearly 80 cents for every marginal dollar they generate. In highly concentrated markets ($HHI > 0.25$), the wage take-home share drops to 63%.

Columns 3 and 4 further divide markets based on whether the self-employment rate is below or above the national average. In less concentrated markets, labor market power remains insignificant regardless of the self-employment rate. However, in highly concentrated markets, the degree of labor market power is influenced by the availability of self-employment oppor-

[23]These figures closely align with those reported by Amodio and De Roux (2022) for Colombian manufacturing plants (inverse elasticity of 0.4) and by Deb et al. (2022) and Yeh, Macaluso and Hershbein (2022) for U.S. manufacturing (ranging from 0.37 to 0.4 and 0.53, respectively). They are slightly lower than what Felix (2022) find for Brazil before the 1990s trade liberalization (50% wage take-home share).

[24]These estimates are derived using more flexible second- and first-stage specifications, where both the log of firm-level employment $\ln l_{i(j,g)t}$ and the instrument $PER_{gt} \times EC_{i(j,g)}$ are interacted with with dummy variables that identify the different subsamples.

[25]This analysis uses contemporaneous wage-bill HHI values. Although potentially endogenous, the contemporaneous HHI determines the wage markdown size at a specific time. When we instrument contemporaneous HHIs with their lags, the resulting pattern closely mirrors the one discussed here. Similarly, when we use previous year's self-employment rates as instruments and categorize markets by self-employment rate, the robustness of the estimates in Columns 3 and 4 is confirmed.

Table 2: Estimates of Labor Market Power

			Self-Employment Rate	
			Low	High
	(1)	(2)	(3)	(4)
All Markets	0.423***			
	(0.052)			
$HHI^{wn} \in (0, 0.18]$		-0.006		
		(0.148)		
$HHI^{wn} \in (0.18, 0.25]$		0.262**		
		(0.105)		
$HHI^{wn} \in (0, 0.25]$			-0.108	-0.061
			(0.087)	(0.128)
$HHI^{wn} \in (0.25, 1]$		0.600***	0.752***	0.104
		(0.150)	(0.112)	(0.067)
SW F-statistics	178.78	222.54	215.76	129.98
		142.77	686.03	725.77
		3369.51		
Observations	6191	6191	3987	2204

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 work 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. Following equation (1), 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.

tunities. The highest level of labor market power is observed in highly concentrated markets with low self-employment rates, where the firm-level inverse labor supply elasticity is estimated at 0.75 and is statistically significant at the 1% level, corresponding to a 57% wage take-home share.[26] In contrast, in highly concentrated markets with higher self-employment rates, the estimated inverse labor supply elasticity is positive but much lower in magnitude and not statistically significant at conventional levels. However, this estimate is statistically different at the 5% level from the corresponding estimate in less concentrated markets. Additionally, the difference between the inverse labor supply elasticity in markets with high versus low self-employment rates among highly concentrated markets is statistically significant at the 1% level, as is the difference-in-differences estimate between highly concentrated and less concentrated markets with varying self-employment rates.

[26]These markets account for a significant portion of manufacturing employment, representing 24% of the manufacturing workforce across all markets for which we have firm-level data. Overall, 66% of all manufacturing workers are in provinces or commuting zones with at least one highly concentrated manufacturing labor market that features low self-employment rates. These areas tend to be less rural and have a higher share of manufacturing employment compared to agriculture.

Discussion A potential concern with our findings is that the local labor market $\times$ year fixed effects in our specification may not fully capture variations in labor supply. For example, if firms facing tighter ex-ante constraints on accessing electricity are located in more rural areas, workers in those areas might respond differently to electrification, potentially violating the exclusion restriction. To address this, we redefine local labor markets as 2-digit industries j within districts d, substantially increasing the granularity of the local labor market $\times$ year fixed effects. This refinement enables us to assess the impact of electrification on labor supply at a geographical level approximately 10 times finer than in the baseline analysis. Results are presented in Online Appendix Table A.9, and all first-stage regression results are reported in Online Appendix Table A.10. The point estimates of the inverse elasticity of labor supply remain broadly consistent with those in Table 2, exhibiting similar patterns of market heterogeneity. Although the estimates are somewhat higher, the standard errors increase as well, reflecting the reduced identifying variation caused by the more granular fixed effects.

Another important consideration is that the *reduced-form* inverse elasticity obtained through our identification strategy does not directly translate into the *structural* inverse elasticity, which measures the elasticity of firm-level wages to employment changes while holding competitors' wages and employment constant. Berger, Herkenhoff and Mongey (2022) address this issue in detail, showing that in granular labor markets, there is no closed-form mapping between the two elasticities; thus, a model is needed to determine the structural, welfare-relevant elasticity. Furthermore, the bias could be either negative or positive depending on the number of treated firms, their market shares, and those of their competitors. Our approach involves replicating the electrification quasi-experiment within the estimated model and comparing the implied reduced-form inverse elasticities with those in Table 2 as a means of testing the model's validity. We then derive the corresponding structural elasticities and rule out the possibility that the observed heterogeneity across markets, related to concentration and self-employment rates, is driven by bias rather than by the underlying structural features of the economy.

The evidence in this section reveal significant interactions between concentration, self-employment opportunities, and wage-setting power. Workers shift between wage employment and self-employment based on earnings. High levels of concentration lead to increased oligopsony power, which results in fewer wage jobs and lower wages, thus pushing more workers towards self-employment. As wages fall, displacing workers from wage employment becomes easier, making workers more responsive to wage changes and, in turn, reducing employers' labor market power. The next section provides a theoretical characterization of these dynamics.

3 Model

We develop a general equilibrium model of Peruvian manufacturing labor markets, where employer concentration, self-employment rates, and labor market power are jointly determined. The model has two primary objectives: (i) to reconcile the empirical evidence presented in the previous section, and (ii) to perform counterfactual policy experiments, particularly in the context of industrial development policies. The assumptions underlying the model are informed by the evidence discussed earlier, as elaborated in Section 3.4.

3.1 Environment

We consider a one-period economy composed of a continuum of local labor markets indexed by $k \in [0, 1]$. Each market consists of a finite number of heterogeneous firms (M_k) and a fixed measure of workers (L_k).

Workers Workers within each market can choose between wage employment in sector F or self-employment in sector S, but they cannot move across different labor markets. In sector F, workers perceive all firms as identical, despite differences in productivity, resulting in a common unit wage. Workers differ in their sector-specific abilities but share identical homothetic preferences for consumption goods with no disutility from labor. Additionally, workers hold equity in firms, meaning their income comes from both labor earnings and profit distributions.

Preferences The numeraire final good is a Cobb–Douglas composite of a continuum of market-level goods:

$$C = \exp\left\{\int_0^1 \alpha_k \log C_k \, dk\right\},$$

where C_k is market-k's variety, and the parameters $\{\alpha_k\}_{k\in[0,1]}$ satisfy $\int_0^1 \alpha_k \, dk = 1$ and determine the shares of income spent on each of these goods.

Each C_k comes in two varieties: $C_{F,k}$, produced by local firms, and $C_{S,k}$, produced by self-employed workers. In turn, each firm $i \in [1, M_k]$ produces a unique variety of $C_{F,k}$. The aggregators are defined as follows:

$$C_k = \left[\zeta C_{F,k}^{\frac{\rho-1}{\rho}} + C_{S,k}^{\frac{\rho-1}{\rho}}\right]^{\frac{\rho}{\rho-1}}, \qquad \text{where} \qquad C_{F,k} = \left(\sum_{i=1}^{M_k} c_{iF,k}^{\frac{\eta-1}{\eta}}\right)^{\frac{\eta}{\eta-1}}.$$

Hence, consumers substitute between $C_{F,k}$ and $C_{S,k}$ with a constant elasticity $\rho > 1$, and among firm-level varieties $\{c_{iF,k}\}_{i\in M_k}$ with a constant elasticity $\eta > 1$. We assume $\eta > \rho$, implying that consumers substitute more easily within sectors than across sectors. Additionally, we introduce a preference shifter for good F, denoted as $\zeta > 0$.

Given this demand structure, each consumer's expenditure on market-level goods is:

$$P_{F,k}C_{F,k} = \gamma_{F,k}\alpha_k Y \qquad \text{and} \qquad P_{S,k}C_{S,k} = (1-\gamma_{F,k})\alpha_k Y, \tag{4}$$

where $P_{F,k} = \left(\sum_{i=1}^{M_k} p_{iF,k}^{1-\eta}\right)^{\frac{1}{1-\eta}}$ is the price index in sector F of market k, and $\gamma_{F,k} = \zeta^{\rho}\left(\frac{P_{F,k}}{P_k}\right)^{1-\rho}$ is the share of expenditure on variety F of good k relative to total expenditure in market k, with

$P_k = \left(\zeta^\rho P_{F,k}^{1-\rho} + P_{S,k}^{1-\rho}\right)^{\frac{1}{1-\rho}}$ being the overall price index in market k.

For individual firm-level expenditure within sector F, demand is:

$$p_{iF,k}c_{iF,k} = s_{iF,k}\gamma_{F,k}\alpha_k Y. \tag{5}$$

Here, $s_{iF,k} = \left(\frac{p_{iF,k}}{P_{F,k}}\right)^{1-\eta}$ is the share of expenditure on variety i of good k in sector F over the total expenditure on sector F in market k.

Labor Supply Workers' skills in each sector are represented by their efficiency units of labor $\mathbf{a}^h \equiv (a_F^h, a_S^h)$, where h denotes a worker. Each worker's ability vector is drawn from a distribution $G_k(a_F, a_S)$, with parameters that may vary across markets.

Let $W_{I,k}$ be the earnings per efficiency unit in sector $J \in \{F, S\}$ of market k. Worker h's earnings in sector J are $E_{J,k}^h = W_{J,k}a_J^h$. Workers take these as given and self-select into wage work or self-employment to maximize earnings. Worker h will choose sector F if and only if:

$$a_F^h W_{F,k} \geq a_S^h W_{S,k} \Leftrightarrow \hat{W}_k \geq \left(\hat{a}^h\right)^{-1},$$

where $\hat{W}_k \equiv \frac{W_{F,k}}{W_{S,k}}$ is the relative wage in market k, and $\hat{a}^h \equiv \frac{a_F^h}{a_S^h}$ is the worker h's relative efficiency, or comparative advantage, in sector F. Therefore, workers with higher $\hat{a}$ have a lower reservation wage for choosing sector F over sector S.

The sorting of heterogeneous workers across sector implies that the aggregate labor supply in sector F can be expressed as the following function of the relative unit wage $\hat{W}_k$:

$$N_{F,k} \equiv N_F(\hat{W}_k) = L_k \int_0^\infty \int_0^{a_F \hat{W}_k} a_F\, g_k(a_F, a_S)\, da_F\, da_S, \quad \text{with} \quad N'_{F,k} > 0. \tag{6}$$

We denote the aggregate elasticity of labor supply as

$$\epsilon_F(\hat{W}_k) \equiv \frac{\partial \ln N_{F,k}}{\partial \ln \hat{W}_k} = \frac{\hat{W}_k \int_0^\infty a_F^2\, g_k(a_F, a_F\hat{W}_k)\, da_F}{\int_0^\infty \int_0^{a_F\hat{W}_k} a_F\, g_k(a_F, a_S)\, da_F\, da_S} > 0.$$

The elasticity of labor supply varies with the relative wage $\hat{W}_k$. As $\hat{W}_k$ shifts, marginal workers either enter or exit wage employment, which alters the composition of the workforce and affects how sensitive the average worker is to wage changes. This dynamic is a key aspect of the labor supply function and will be essential for understanding labor market power in the economy.

Technology Production technology in sectors S and F is linear in efficiency units of labor. Total output in sector S is given by:

$$Y_{S,k} = N_{S,k}, \tag{7}$$

where $N_{S,k}$ is the efficiency labor units in sector S. In sector F, the output of firm i is:

$$y_{iF,k} = z_{iF,k} n_{iF,k}, \quad \forall\, i = \{1, \ldots, M_k\},$$

where $z_{iF,k}$ is the firm's idiosyncratic productivity and $n_{iF,k}$ is the firm's labor demand.

Let $Y_{F,k} = \left(\sum_{i=1}^{M_k} y_{iF,k}^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}}$ denote aggregate output in sector F. We can then write:

$$Y_{F,k} = Z_{F,k} N_{F,k}, \tag{8}$$

where $N_{F,k} \equiv \sum_{i=1}^{M_k} n_{iF,k}$ is aggregate labor demand 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 a productivity index for sector F of market k. Labor market clearing requires that $N_{F,k}$ in equation (8) equals aggregate labor supply in equation (6), while $N_{S,k}$ in equation (7) equals aggregate labor supply in sector S, which can be derived analogously to equation (6).

Market Structure Sector S operates in perfectly competitive labor and output markets, where goods are sold at marginal cost. In contrast, firms in sector F engage in Nash-Cournot competition in both product and labor markets. In product markets, firms produce differentiated varieties of the final good, leading to heterogeneous markups and prices. However, in the labor market, firms are viewed as perfect substitutes, resulting in oligopsonistic competition among essentially homogeneous employers, which leads to a common unit wage.

Firm's Problem Each firm i chooses its labor demand to maximize profits, subject to demand and aggregate labor supply functions. Firms take aggregate prices P_k and $P_{S,k}$ as given, but internalize the effect of their labor demand on the aggregate price $P_{F,k}$ and wage $W_{F,k}$.

Solving for the firm-level equilibrium yields the first-order condition:

$$W_{F,k} = \frac{MRPL_{iF,k}}{\psi_{iF,k}}, \tag{9}$$

where $MRPL_{iF,k} \equiv \frac{\partial R_{F,k}}{\partial N_{F,k}} = \frac{p_{iF,k} z_{iF,k}}{\mu_{iF,k}}$ is the firm's marginal revenue product of labor (MRPL). The wage markdown $\psi_{iF,k}$ of each firm i, which is given by:

$$\psi_{iF,k} = 1 + \frac{s_{iF,k}^{N}}{\epsilon_F(\hat{W}_k)} \geq 1, \tag{10}$$

measures the firm's labor market power. When it exceeds 1, the market wage is below firm's MRPL, indicating wage-setting power. The markdown increases with the firm's employment share $s_{iF,k}^{N} \equiv \frac{n_{iF,k}}{N_{F,k}}$, while it decreases with the market-level labor supply elasticity.[27]

[27]Specifically, the markdown is defined as: $\psi_{iF,k} \equiv 1 + \frac{1}{\epsilon_{iF,k}}$, where $\epsilon_{iF,k}$ denotes firm i's residual elasticity of labor supply, holding fixed the employment at other firms: $\epsilon_{iF,k} \equiv \left. \frac{\partial \ln n_{iF,k}}{\partial \ln W_{F,k}} \right|_{n_{-iF,k}}$. This residual elasticity can

Substituting the expression for $MRPL_{iF,k}$ into (9), we can express the firm-level price as a markup over marginal cost:

$$p_{iF,k} = \mu_{iF,k}\psi_{iF,k}\frac{W_{F,k}}{z_{iF,k}}. \tag{11}$$

The markup term $\mu_{iF,k}$ in equation (11) is defined as:

$$\mu_{iF,k} = \frac{\varepsilon_{iF,k}}{\varepsilon_{iF,k}-1}, \quad \text{where} \quad \varepsilon_{iF,k} \equiv \varepsilon(s_{iF,k}) = \left[\frac{1}{\eta}(1-s_{iF,k}) + \frac{1}{\rho}s_{iF,k}\right]^{-1}, \tag{12}$$

and captures the product market power of firm i. As in standard models with oligopolistic competition, it depends on the demand elasticity $\varepsilon_{iF,k} \in [\rho, \eta]$, which is a function of the firm's market share $s_{iF,k}$. The demand elasticity decreases—and hence the markup $\mu_{iF,k}$ increases—as the firm's market share grows.

3.2 Equilibrium

In our model of segmented labor markets, interactions across markets occur only through changes in expenditures $Y_k \equiv \alpha_k Y$, where $\{\alpha_k\}_{k\in(0,1)}$ are the constant expenditure shares. Given Y, the equilibrium in each market can be solved independently of the others.

This structure allows to facilitate the solution into two components: market equilibrium and general equilibrium. Below, we briefly outline these components. The detailed algorithm and numerical implementation can be found in Online Appendix B.1.

Market Equilibrium The market equilibrium refers to the process of solving for equilibrium in each local labor market given Y and model's fundamentals. It is characterized by a vector $\hat{\mathbf{K}}_k \equiv \left\{M_k, \hat{W}_k, \mathbf{\Lambda}_k\right\}$ for each market k — where M_k is the number of active firms, $\hat{W}_k$ is the relative wage, and $\mathbf{\Lambda}_k = \{s_{iF,k}, s^N_{iF,k}, \mu_{iF,k}, \psi_{iF,k}\}_{i=1}^{M_k}$ represents the vector of output and employment shares, markups, and markdowns for each firm — that satisfy equations (4)-(12) for given Y and $\{\alpha_k, \{z_{iF,k}\}_{i\in[1,M_k]}, G_k, f^e_k\}_{k\in(0,1)}$.

We first assume that the set of employers M_k and their productivity is known for each k. Given a guess for $\hat{W}_k$, equations (5), (10), and (11) define a fixed-point problem that can be solved for the vector $\mathbf{\Lambda}_k$. In turn, given the vector $\mathbf{\Lambda}_k$, the relative wage $\hat{W}_k$ can be found from equations (4), (7), and (8). The resulting fixed point in each market k constitutes the vector of market equilibria, given a guess for M_k.

be expressed as:

$$\epsilon_{iF,k} = \frac{\partial \ln n_{iF,k}}{\partial \ln N_{F,k}} \cdot \frac{\partial \ln N_{F,k}}{\partial \ln W_{F,k}} = \frac{\epsilon_{F,k}}{s^N_{iF,k}},$$

where $\epsilon_{F,k} \equiv \frac{\partial \ln N_{F,k}}{\partial \ln W_{F,k}}$ represents the market-level elasticity of labor supply, and $\frac{\partial \ln n_{iF,k}}{\partial \ln N_{F,k}} = \frac{1}{s^N_{iF,k}}$, which follows from the identity $N_{F,k} = \sum_{i=1}^{M_k} n_{iF,k}$.

Solving for the Number of Entrants We assume that, upon entry, firms must pay a fixed cost $f^e_{i,k}$ in units of the final good. Entry is modeled as a sequential game, with more productive firms entering first. The process involves iteratively solving for equilibrium wages and market shares using the fixed-point algorithm above. The profits of the marginal entrant are then calculated, and equilibrium is achieved when these profits are non-negative, while any additional entrants would incur losses. This results in a unique and stable cutoff equilibrium, where only firms above a certain productivity threshold enter.

To reduce computational intensity, particularly when solving the fixed-point algorithm for each candidate M_k, we adopt a simplified entry model for baseline calibration, following Gaubert and Itskhoki (2021). In this approach, firms are assumed to behave "naively" at the entry stage, expecting atomistic markups and markdowns, which makes the computation of market shares and equilibrium conditions more tractable. For robustness, we later show that the full entry game produces similar results.[28]

General Equilibrium The general equilibrium is defined by a vector of income and prices $\mathbf{X} = (Y, P)$, with $P = 1$ by normalization, such that aggregate income equals expenditure, and product markets clear. Aggregate income is:

$$Y = \int_{k \in (0,1)} \left[E_k + \Pi_{F,k} + F^e_k\right] dk, \tag{13}$$

where the three terms on the right-hand side correspond to (i) total labor income in market k, $E_k \equiv W_{S,k} N_S(\hat{W}_k) + W_{F,k} N_F(\hat{W}_k)$, (ii) aggregate firm profits $\Pi_{F,k} = \sum_{i \in [1,M_K]} \pi_{iF,k}$, and (iii) total entry cost $F^e_k \equiv \sum_{i \in [1,M_K]} f^e_{i,k}$. Product market clearing requires that:

$$Y = C. \tag{14}$$

Given the market equilibrium $\mathbf{K} = \left\{M_k, \hat{W}_k, \mathbf{\Lambda}_k\right\}_{k \in (0,1)}$, the general equilibrium $\mathbf{X}$ solves equations (13)-(14). Conditional on $\mathbf{X}$, the market equilibrium and entry game yield the market equilibrium $\mathbf{K}$. The fixed point $(\mathbf{X}; \mathbf{K})$ is the economy equilibrium.

3.3 Characterization

Table 2 shows that labor market power rises with employer concentration, but self-employment weakens this relationship. However, because the reduced-form elasticities capture both direct and equilibrium effects, identifying the precise nature of these patterns is challenging. We argue that these co-movements are intrinsic to the structural features of our model economy.

Let $\bar{\psi}_{F,k} \equiv \sum_{i \in M_k} s^N_i \psi_{iF,k}$ denote the weighted average of firm-level markdowns in market

[28] See Section 4.5.3 for details.

k, with employment shares as weights. From equation (10), we can express this as:

$$\bar{\psi}_{F,k} = 1 + \frac{HHI^{wb}_{F,k}}{\epsilon_F(\hat{W}_k)}, \tag{15}$$

where $HHI^{wb}_{F,k}$ is the wage-bill HHI in sector F of market k, which in our model coincides with the employment-based HHI. This equation shows that average labor market power in a local labor market increases with employer concentration, but the effect is weaker when the aggregate labor supply elasticity $\epsilon_F(\hat{W}_k)$ is high.

Equation (15) shows that the endogenous adjustment of $\epsilon_F(\hat{W}_k)$ to wage changes are crucial for understanding labor market power. The nonparametric characterization of these adjustments, however, is intractable. We make progress by imposing parametric assumptions on the distribution of workers' abilities, specifically focusing on the log-normal case. This approach is commonly used in empirical Roy models due to its favorable identification properties, which we will discuss further in Section 4.2.

When workers' abilities follows a log-normal distribution, and under typical parameterizations found in empirical studies consistent with our data, $\epsilon_F(\hat{W}_k)$ can be approximated as:

$$\epsilon_F(\hat{W}) \approx \frac{\lambda(c_F)}{\sigma^*}, \qquad \text{with} \qquad c_{F,k} = \frac{\ln \hat{W}_k + \hat{\mu}}{\sigma^*}, \tag{16}$$

where $\sigma^* > 0$ and $\lambda(x) = \frac{\phi(x)}{\Phi(x)}$ is the ratio of the standard normal probability density function $\phi(x)$ to the standard normal cumulative distribution function $\Phi(x)$, which implies $\lambda'(\cdot) < 0$.[29]

This result has important implications. First, it shows that $\epsilon_F(\hat{W})$ decreases monotonically with $\hat{W}_k$. Second, it says that the elasticity depends only on the variance-covariance parameters of the ability distribution and the difference in population means. In Section 4.2, we will show that these parameters can be identified from cross-sectional worker earnings data in the log-normal case, enabling us to trace out $\epsilon_F(\cdot)$ and its effect on labor market power.

Additionally, in the log-normal case, $c_{F,k}$ is a monotonically decreasing function of the self-employment share (self rate$_k$), i.e.,

$$c_{F,k} = \Phi^{-1}(\text{1-self rate}_k). \tag{17}$$

Hence, the self-employment share is a sufficient statistic for the labor supply elasticity, which increases with the self-employment share. Equations (15)-(17) capture the core insight of our theory. Concentration increases average markdowns via oligopsony power, reducing wages. As a result, more workers opt for self-employment. This shift alters the composition of workers in both sectors, increasing labor supply elasticity and reducing labor market power.

[29]The exact expression is: $\epsilon_F(\hat{W}) = \frac{1}{\sigma^*}\left(\lambda(c_F) + \alpha\lambda'(c_F) + \frac{\alpha}{2}\lambda''(c_F)\right)$, where $\alpha = \frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}$, and $\sigma^* = \sqrt{\sigma_F^2 + \sigma_S^2 - 2\varrho\sigma_F\sigma_S}$. The approximation holds because α is typically small, allowing us to approximate $\alpha \approx 0$. For instance, our estimates imply a value of $\alpha = 0.0002$. See Online Appendix B.2.1 for a detailed discussion.

The Pass-Through of Shocks to Average Earnings Our theory also delivers insights into the effects of shocks to the economic environment on labor market outcomes. For illustration, we focus on average sectoral earnings, a key policy outcome in the counterfactuals below. In sector F, we can decompose the change in (log) average earnings, $\ln \bar{E}_{F,k}$, as:

$$d \ln \bar{E}_{F,k} = - \underbrace{d \ln \bar{\psi}_{F,k}}_{\text{labor market power}} + \underbrace{d \ln Z_{F,k} + d \ln P_{F,k} - d \ln \bar{\mu}_{F,k}}_{\text{labor revenue productivity } (\overline{MRPL}_{F,k})} + \underbrace{d \ln \bar{A}_{F,k}}_{\text{selection}}, \tag{18}$$

where $P_{F,k}$ is the sectoral price index and $\bar{\mu}_{F,k}$ is the average markup, defined as $\bar{\mu}_{F,k} \equiv \left(\sum_i s_{iF,k} \cdot (\mu_{iF,k})^{-1}\right)^{-1}$.

There are two main channels through which shocks affect average earnings in sector F: the *direct* effect on the wage per efficiency unit, captured by the first two bracket terms, and the *selection* effect, which reflects how the average worker ability in the sector changes in response to the shock, and depends on the assumptions on the distribution of worker abilities. The direct effect is itself divided into two components. The *labor market power* channel captures how the average markdown changes with the shock. The *labor revenue productivity* channel reflects how the average marginal revenue product of labor changes with the shock, through changes in aggregate productivity ($Z_{F,k}$), prices ($P_{F,k}$), and average markups ($\bar{\mu}_{F,k}$).

Changes in earnings in the self-employment sector can be decomposed in a similar fashion, but with perfect competition, the expression simplifies to:

$$d \ln \bar{E}_{S,k} = \underbrace{d \ln P_{S,k}}_{\overline{MRPL}_{S,k}} + \underbrace{d \ln \bar{A}_{S,k}}_{\text{selection}}. \tag{19}$$

The sorting of heterogeneous workers across sectors significantly influences how shocks affect average sectoral earnings by altering the average ability of workers in each sector, i.e., through the *selection* channel. These effects depend on the correlation between workers' abilities in wage work and self-employment, as well as the relative dispersion of these abilities (Adão, 2016; Amodio, Alvarez-Cuadrado and Poschke, 2020). In Online Appendix B.2.2, we show that, when the two abilities are strongly positively correlated and more dispersed in self-employment than in wage work, the mean ability of workers in both sectors decreases (increases) if the relative wage $\hat{W}_k$ falls (rises) in response to the shock. This is because absolute and comparative advantage are negatively correlated in wage work but positively correlated in self-employment, which implies that as more workers select into (out of) wage employment, the average ability increases everywhere. Vice versa, if abilities are negatively correlated or if the correlation is positive but low, the mean ability of wage workers will increase, and the one of self-employed workers will decrease as more workers choose self-employment.

3.4 Model Discussion

We conclude with a discussion of four key assumptions in our model. First, we assume that workers perceive all firms in sector F as homogeneous, enabling us to model the labor market as a standard Cournot oligopsony. Firms strategically choose labor demand but offer a uniform wage, despite productivity differences. This simplifies our model compared to recent oligopsony theories where firms are imperfect substitutes, leading to wage variation.[30] This choice is motivated by two factors: (i) it ensures analytical tractability, allowing us to explore the sources of market-level differences in labor market power, which is central to both the theoretical and empirical analysis; (ii) it is essential for model estimation, as discussed next. However, unlike the related literature, Cournot competition is the only form of Nash conduct that leads to markdowns in our model, whereas Bertrand competition would result in efficient outcomes.

The second key assumption is that we restrict worker mobility across markets, effectively segmenting labor markets. This assumption is supported by the evidence in Section 2 showing that transitions between wage work and self-employment are more frequent than movements between local labor markets within wage work, a common focus in the literature. The key advantage of this assumption is that it simplifies worker sorting into a binary decision, similar to the classic Roy (1951) model. Together with our assumptions about labor market structure and the log-normal parameterization of the ability distribution, this facilitates the mapping of the ability distribution to cross-sectional earnings data for identification. As discussed in Section 3.3, an important implication is that we can use worker-level data to discipline the labor supply determinants of labor market power in the model.

Finally, a notable feature of our model is that it incorporates oligopoly power in the output market, oligopsony power in the labor markets, and endogenous entry. This sets our approach apart from much of the existing literature, which often assumes fixed entry due to the computational challenges of modeling entry games with oligopsony. Key difficulties in modeling endogenous entry include accounting for existing competitors, which can lead to multiple equilibria, and determining entry patterns across interdependent markets (MacKenzie, 2021).

Two key assumptions make the entry problem tractable in our model. First, the boundaries of the product and labor markets align, with product varieties and worker decisions being determined within both industry and location.[31] Second, we assume Cobb-Douglas preferences across market-level goods. Combined with segmented labor markets, these assumptions ensure that a firm's market share depends only on local competitors, allowing firms to make independent entry decisions across different markets.

[30] See, e.g., MacKenzie (2021); Berger, Herkenhoff and Mongey (2022); Felix (2022); Gutiérrez (2023).

[31] While this is common for labor markets, it is less typical for product markets, which are often defined at the national level. Berger, Herkenhoff and Mongey (2022) assume perfect competition in the output market. Gutiérrez (2023) discusses the challenges that arise when product and labor market boundaries do not align, in a theory of oligopoly and oligopsony with fixed entry.

4 Model Identification and Estimation

This section explains how the model is identified and estimated using Peruvian data. First, we outline the parameterization of the model. Second, we describe how we identify the parameters of the joint ability distribution using direct inference methods. Third, we discuss the joint estimation of all remaining parameters using the method of simulated moments. Finally, we evaluate the model fit of both targeted and untargeted moments.

4.1 Parameterization

We adopt a parameterization of the model that incorporates heterogeneity across local labor markets in key factors influencing firms' and workers' decisions, notably firm productivity, entry costs, and worker ability. Specifically, we consider each local labor market as a multi-dimensional observation from the structural data-generating process outlined by the model, with common parameters to be estimated.[32]

Firm Productivity Firm productivity may be influenced by both idiosyncratic factors and characteristics related to the local labor market. We consider a parameterization of productivity that accounts for both sets of factors.

Let M_k^* denote the potential (shadow) entrants in the wage sector of local labor market k. We assume that M_k^* follows a Poisson distribution with $\mathbb{E}(M_k^*) = \bar{M}_k^*$. Each potential entrant draws an i.i.d. productivity value from a Pareto distribution with lower bound $\underline{z}_k$ and shape parameter θ, with lower values of θ indicating a more dispersed and skewed distribution. Under this Poisson-Pareto structure, the parameter $T_k \equiv \bar{M}_k^* \cdot \underline{z}_k^\theta$ emerges as a sufficient statistic for determining expected productivity in a local labor market.[33] This structure of productivity draws, inspired by Gaubert and Itskhoki (2021), not only provides a tractable modeling environment but also generates a realistic cross-sectional distribution of firm sales.

To account for the heterogeneity in productivity across local labor markets, we then assume that the T_k values are drawn from a log-normal distribution with parameters μ_T and σ_T:

$$T_k \sim \log \mathcal{N}(\mu_T, \sigma_T).$$

This assumption is motivated by the evidence in Online Appendix Figure A.11, which shows that the distribution of log sales across local labor markets is well approximated by a log-normal distribution.

[32]This approach to modeling across-market heterogeneity contrasts with a more direct model inversion, where each local labor market (firm) in the data is directly mapped to a corresponding market (firm) in the model (See, e.g., Gutiérrez (2023)). Our model's structure does not facilitate straightforward inversion, making the direct estimation of parameters from observable data more challenging than in other work.

[33]Specifically, the number of shadow firms with productivity above $z > \underline{z}_k$ follows a Poisson distribution with mean $T_k z^{-\theta}$. As long as the least efficient firm remains inactive, the model's predictions remain invariant to different combinations of $\bar{M}_k^*$ and $\underline{z}_k$ that yield the same T_k (Eaton, Kortum and Sotelo, 2012).

Firm Entry Cost It is common to parameterize entry costs as a constant across firms. For our baseline specification, we depart from this approach with two key adjustments. First, to ensure that at least one firm operates in each local labor market, we set the entry cost for the first firm to zero, $f^e(1) = 0$. Second, for subsequent entrants, we allow entry costs to increase with the number of active firms.[34] Specifically, for entrants ranked $n \geq 2$, the entry cost is defined as:

$$f^e(n) = f_0 + f_1\sqrt{n}, \quad \text{for } n \geq 2.$$

This fixed cost structure allows us to successfully replicate key moments in the distribution of entrants across markets. In Section 4.5, we consider an alternative parameterization where f^e_k varies across local labor markets.

Worker Ability Lastly, we assume that each worker's endowment of efficiency units of labor in the two sectors, $\mathbf{a} = (a_F, a_S)$, follows a joint log-normal distribution:

$$\log \mathbf{a} \sim \mathcal{N}\left(\boldsymbol{\mu_k}, \boldsymbol{\Sigma_k}\right), \quad \text{where} \quad \boldsymbol{\mu_k} = \begin{pmatrix} \mu_{F,k} \\ \mu_{S,k} \end{pmatrix}, \quad \boldsymbol{\Sigma_k} = \begin{pmatrix} \sigma^2_{F,k} & \varrho_k \sigma_{F,k}\sigma_{S,k} \\ \varrho_k \sigma_{F,k}\sigma_{S,k} & \sigma^2_{S,k} \end{pmatrix}. \tag{20}$$

Here, $\boldsymbol{\mu_k}$ represents the mean abilities in wage work and self-employment, while $\boldsymbol{\Sigma_k}$ captures the variance-covariance structure that governs comparative advantage.

The log-normal assumption for G_k is common in empirical Roy models as it facilitates the identification of the underlying parameters (French and Taber, 2011). In Section 4.2.1, we explain how, within our framework, most parameters can be identified on a market-by-market basis using cross-sectional earnings data from both wage and self-employment sectors. Specifically, we can identify the variance-covariance parameters $\boldsymbol{\Sigma_k}$ and the relative comparative advantage $\hat{\mu}_k \equiv \mu_{F,k} - \mu_{S,k}$ at the market level. However, the absolute advantage parameters $\mu_{F,k}$ and $\mu_{S,k}$ cannot be directly identified. To address this, we employ an indirect approach by assuming that $\mu_{S,k}$ is drawn from a normal distribution:

$$\mu_{S,k} \sim \mathcal{N}\left(\mu_{\mu_S}, \sigma_{\mu_S}\right),$$

where the parameters μ_{μ_S} and σ_{μ_S} are estimated using the MSM procedure described in Section 4.2.2. Given a draw for $\mu_{S,k}$ and an estimate for $\hat{\mu}_k$, we can then recover $\mu_{F,k}$ as $\mu_{F,k} = \mu_{S,k} + \hat{\mu}_k$.

The assumption of normality for $\mu_{S,k}$ is supported by the evidence in Online Appendix Figure A.11 showing that the distribution of years of education across local labor markets is well approximated by a normal distribution. Later, we will show that this parameterization of the ability distribution yields a plausible distribution of workers' abilities and earnings.

[34] This adjustment is crucial to capture the high concentration observed in Peruvian labor markets and to ensure that the last shadow entrant always remains inactive. This is necessary for accurately reflecting average market productivity through the parameter T_k, as discussed earlier.

4.2 Estimation Strategy

With these parametric assumptions, we estimate the model in three steps. First, we calibrate the Cobb-Douglas expenditure shares $\{\alpha_k\}_{k\in[0,1]}$ and population shares $\{L_k\}_{k\in[0,1]}$ from the data, equating them to the income share and the share of the workforce in each local labor market.[35]

In the second step, described in Section 4.2.1, we use our matched firm-worker level data to estimate the parameters of the variance-covariance matrix of the workers' ability distribution ($\boldsymbol{\Sigma_k}$) as well as the mean comparative advantage ($\hat{\mu}_k$). In the third and final step, detailed in Section 4.2.2, we implement a MSM procedure to estimate all the remaining parameters.

4.2.1 Direct Inference

We briefly outline the strategy for estimating the parameters of the workers' ability distribution, $\boldsymbol{\Sigma_k}$ and $\hat{\mu}_k$. Online Appendix C.1 provides a detailed explanation of our approach.

Variance-Covariance Parameters Identification of $\boldsymbol{\Sigma_k} = \{\sigma_{F,k}, \sigma_{S,k}, \varrho_k\}$ builds on the approach in Heckman and Sedlacek (1985), and relies on the relationship between observed labor market outcomes and the underlying ability distribution of workers. Under the assumption of log-normality, these relationships are straightforward to express. By analyzing worker shares and the mean and variance of their earnings in each sector, we can obtain consistent estimates of the $\boldsymbol{\Sigma_k}$ parameters for each market.

Mean Comparative Advantage The mean comparative advantage $\hat{\mu}_k$ governs the relationship between the average ability gap across sectors, the share of workers in each sector, and the parameters $\boldsymbol{\Sigma_k}$. The lack of data on abilities poses an identification challenge. To address this, we employ an identification strategy that uses years of education as a proxy for unobserved abilities. By combining this proxy with the structural model equations, we estimate the ability gap between workers from the observed average education gap.[36]

We define β as the elasticity of relative ability with respect to relative education and assume that this elasticity is constant across markets. This parameter captures how differences in average (log) education levels between wage and self-employed workers translate into differences in average (log) abilities across sectors. We estimate β by analyzing the relative differences in education and earnings across sectors within each market, deriving an empirical regression

[35] For the model calibration, we use a merged sample comprising local labor markets where both firm-level and worker-level data across both sectors are available. Summary statistics for this sample are provided in Online Appendix Table A.1. The sample includes 1,040 local labor market-year observations. For computational efficiency, we reduce the sample size to 234 for the baseline calibration. A histogram of the resulting $\{\alpha_k, L_k\}$ is presented in Online Appendix Figure A.10.

[36] In Heckman and Sedlacek (1985), a similar identification challenge for the absolute advantage parameters μ_F and μ_S is addressed using a standard instrumental variables approach with time-series wage data. In contrast, our approach to estimating the remaining parameters combines direct and indirect methods, consistent with our modeling framework.

equation from the model's structural equations. With estimates for β and $\boldsymbol{\Sigma_k}$, we then infer $\hat{\mu}_k$ for each market based on the expression for mean log ability gap when abilities are log-normal.

Setting Parameters to Constant Despite our ability to recover market-specific parameters, we maintain a constant variance-covariance matrix and mean comparative advantage across markets in our baseline model for two main reasons. First, this parsimonious approach enhances transparency by minimizing market heterogeneity and the impact of measurement error in earnings data. Second, it aligns with our method of matching the model to the data, where the mapping between markets in the data and in the model is indirect. In Section 4.5, we discuss an extension of the model that allows for heterogeneity in these parameters.

4.2.2 Targeted Moments and Identification of Remaining Parameters

We now outline the procedure used to estimate the remaining parameter vector as given by $\Phi = (\mu_T, \sigma_T, \theta, f_0, f_1, \mu_{\mu_S}, \sigma_{\mu_S}, \eta, \rho, \zeta)$. We target 27 empirical moments that capture key local labor market outcomes, informed by the empirical results in Section 2. These moments reflect the cross-sectional characteristics of concentration, self-employment, and their co-movements. While any parameter variation influences all moments, certain parameters are more directly related to specific moments. Below, we discuss the key factors driving identification.

Targeted Moments First, we focus on capturing the prevalence of concentration across local labor markets. From Table 1, we target the mean and standard deviation of the (log) number of firms in each market, as well as the employment-based Herfindahl-Hirschman Index (HHI), both weighted and unweighted. We also target the share of local labor markets with a single employer and the corresponding percentage of total wage employment.[37]

These moments are essential for identifying productivity parameters and the fixed cost distribution. Intuitively, fixed cost parameters (f_0 and f_1), along with productivity parameters for firms and workers (θ and μ_{μ_S}), determine the incidence of concentration across markets. A market can feature fewer firms either due to high fixed costs or high firm or worker productivity.

We also target moments of the distribution of log total sales across markets, specifically the interquartile ratio and the 90-10 ratio, along with the mean and standard deviation of the sales concentration ratio (CR1 and CR4) within markets. These moments help identify the average market productivity parameters (μ_T and σ_T) and within-market productivity variation (θ).

Next, we aim to capture self-employment patterns across local labor markets. We target the mean and standard deviation of the wage employment share and the relative (log) worker earnings between wage and self-employed workers, as reported in Table 1. Under our parameterization of the ability distribution, with fixed $\boldsymbol{\Sigma}$ and $\hat{\mu}$, the wage employment share is a

[37]We specifically target the moments presented in Online Appendix Table A.1, which replicates Table 1 for the merged sample of local labor markets where both firm-level and worker-level data are available across sectors.

monotonic function of the relative wage $\hat{W}$, which depends on the elasticity parameter ρ and the preference shifter ζ. These two parameters also influence the mean (log) relative earnings.

Although most ability distribution parameters are estimated externally, we target moments of the relative average schooling between wage and self-employed workers—mapped to relative average ability in the model—focusing on the interquartile and 90-10 ratios. These moments help identify the mean absolute advantage parameters $(\mu_{\mu_S}, \sigma_{\mu_S})$, which determine the dispersion in average ability across markets.

Finally, we target correlations between the employment-based HHI and several labor market outcomes, specifically the wage-employment share and (log) earnings in both sectors. The sensitivity of these variables to changes in employer concentration mainly depends on the across-sector elasticity (ρ) and the absolute advantage parameters $(\mu_{\mu_S}, \sigma_{\mu_S})$.

Normalization To improve estimation precision, we apply the following normalizations. First, since the elasticity of substitution η is weakly separately identified from the productivity parameter θ, both being linked to the Pareto tail of the sales distribution across firms, we fix $\eta = 6$ and estimate θ in the MSM routine. Second, as the parameter f_1 shows minimal sensitivity to the targeted moments, we set it externally to 5×10^{-7}. This reduces the parameter vector to $\Phi = (\mu_T, \sigma_T, \theta, f_0, \mu_{\mu_S}, \sigma_{\mu_S}, \rho, \zeta)$, thereby improving estimation precision.

Identification To support our identification strategy, we formally examine the connection between parameters and moments by computing the elasticity of each model-generated moment with respect to each parameter, following standard practices in the literature (e.g., Kaboski and Townsend 2011). The full Jacobian matrix is provided in Online Appendix Figure A.13. Below, we offer some insights based on the results of this analysis.

We find that the parameters θ, ζ, and ρ are the most influential for the majority of the targeted moments. This is expected, as these parameters play key roles in determining the equilibrium relative wage $\hat{W}_k$ in each local labor market, as shown in equation (A.3) in the Online Appendix B.1. By targeting multiple moments directly linked to the relative wage, more than the number of unknown parameters, we ensure sufficient identifying variation for these parameters. Additionally, only the moments related to relative ability show sensitivity to changes in the dispersion parameter for workers' absolute advantages. This is reassuring, as the market equilibrium—which determines most (other) moments—depends solely on $\hat{\mu}$, not on the absolute advantage parameters.

Overall, the Jacobian matrix confirms the identification argument, demonstrating that the model provides enough variation to identify the remaining parameters effectively.

Table 3: Summary of Model Parameters

Parameter		Value
Panel I. Externally Fixed		
η	Substitution elasticity within sector F	6
f_1	Fixed cost slope parameter	5×10^{-7}
Panel II. Externally Estimated		
σ_F	St. dev. of log ability as a wage worker	0.81
σ_S	St. dev. of log ability as a self-employed	0.91
ϱ	Correlation of log abilities	0.89
$\hat{\mu}$	Mean comparative advantage	-0.12
Panel III. Estimated via MSM		
μ_T	Mean of market-level productivity	0.82
σ_T	St. dev. of market-level productivity	0.96
f_0	Fixed cost intercept parameter	1.97×10^{-3}
θ	Firm-level productivity dispersion parameter	2.34
μ_μ	Mean of market-level mean absolute advantage	1.68
σ_μ	St. dev. of market-level mean absolute advantage	0.10
ρ	Substitution elasticity across F and S	2.72
ζ	Sector F preference shifter	1.88

Notes. This table reports the parameter values for the quantitative model. See Section 4 for details on parameter estimation.

4.3 Estimation Results

Table 3 summarizes the estimated parameter vector. Panel II provides the median estimates of the key Roy model parameters, based on the direct inference approach outlined in Section 4.2.1. Online Appendix Figure A.12 displays the histograms of the estimated variance-covariance parameters and mean comparative advantage across markets.

The two abilities are highly correlated, with an estimated correlation coefficient of $\hat{\varrho} = 0.89$, and the ability for self-employment is more dispersed than the ability for wage work, i.e., $\hat{\sigma}_S > \hat{\sigma}_F$. These parameters are precisely estimated, with bootstrap standard errors ranging from 0.02 to 0.07. The estimates suggest no correlation between workers' comparative and absolute advantage in wage employment, but a positive correlation in self-employment advantages. This is consistent with the evidence in Figure 1, which shows that transitions into and out of wage employment are more common among lower-earning self-employed workers, while transitions to self-employment are unrelated to wage earnings.[38] We also estimate $\hat{\mu} = -0.12$, indicating that the average worker in the population has a comparative advantage in

[38]These transitions can be explained by shocks to relative unit earnings $\hat{W} = W_F/W_S$ combined with the estimates of the variance-covariance matrix of the joint ability distribution. Positive selection in self-employment implies that transitions are more frequent among lower-earning self-employed workers. Meanwhile, the lack of selection in wage work suggests that sector switchers earn wages comparable to those of inframarginal wage workers. Changes in $\hat{W}$ and their heterogeneity in terms of sign, size, and frequency across markets, combined with the estimated sign and strength of selection, are therefore sufficient to generate the patterns in Figure 1.

Table 4: Targeted Moments and Model Fit

Moment	Model	Data	Moment	Model	Data
Panel I. Distribution Moments					
Log Number of Firms			Wage-bill Share of		
Mean	0.97	1.22	Markets with 1 firm	0.07	0.08
Standard Deviation	0.95	1.17	Markets with $<$10 firms	0.89	0.84
			Markets with $<$50 firms	1.00	0.99
Log of Sales					
Ratio p75/p25	2.94	2.92	Share of Wage Employment		
Ratio p90/p10	5.29	5.30	Mean	0.66	0.71
CR_1, Mean	0.66	0.69	Standard Deviation	0.12	0.32
CR_1, Standard Deviation	0.29	0.29			
CR_4, Mean	0.94	0.91	Log of Earnings$_F$/Earnings$_S$		
CR_4, Standard Deviation	0.11	0.15	Mean	0.41	0.40
			Standard Deviation	0.58	0.93
Employment HHI					
Mean, Unweighted	0.57	0.59	Log of Schooling$_F$/Schooling$_S$ (Ability$_F$/Ability$_S$)		
Standard Deviation	0.34	0.35	Ratio p75/p25	1.42	1.28
Mean, Weighted	0.30	0.33	Ratio p90/p10	1.18	1.04
Percent of Markets with 1 firm	0.36	0.39			
Panel II. Regression Coefficients					
% Wage Employment on (Log) HHI^{wb}			(Log) Earnings$_F$ on (Log) HHI^{wb}		
Point Estimate	-0.04	-0.07	Point Estimate	-1.36	-0.13
Standard Error	0.01	0.01	Standard Error	0.13	0.02
(Log) Earnings$_S$ on (Log) HHI^{wb}					
Point Estimate	-1.17	-0.11			
Standard Error	0.12	0.03			

Notes. This table reports the moments used in the estimation and compares them with those calculated from the estimated model. The data moments are computed in the sample of local labor markets where at least one formal firm is active and the share of self-employed workers and wage workers is strictly between 0 and 1. See Section 4 for more details on the moments' construction.

self-employment. These findings align with experimental evidence showing that workers tend to prefer self-employment over industrial jobs in poor countries (Blattman and Dercon, 2018).

Panel III presents the estimated parameter vector from the MSM procedure, with the corresponding model moments summarized in Table 4. The model demonstrates a strong fit to the data, which is noteworthy given that only 8 parameters were used to target 27 moments.

The model effectively captures various measures of concentration across local labor markets. It closely replicates the high share of monopsonistic labor markets, with 39% observed in the data and 36% predicted by the model, as well as the corresponding payroll share–8% in the data and 7% in the model. Additionally, the model predicts that approximately 66% of workers are wage employees, compared to 71% in the data. Wage workers in the model earn about 0.4 log points more than self-employed workers, in line with the evidence.

The model also successfully replicates the negative cross-sectional correlations between earnings in both sectors and wage-employment rates with the payroll-based HHI. All coefficients are both economically and statistically significant. However, the model falls short in accurately matching the correlations between concentration and mean log earnings in the two sectors. Although the signs of the coefficients are correct, their magnitudes deviate from the observed data. We find that this discrepancy is mostly driven by markets with one firm, to which the model assigns lower average earnings than in the data.[39]

Overall, the model's ability to replicate the core patterns documented in Section 2 builds confidence in our estimated parameters, which are broadly consistent with findings from other studies in the literature. For instance, we estimate a Pareto shape parameter of $\theta = 2.34$, which is higher but close to the 1.5 in Huang et al. (2024) for Chilean importers.[40] We also find limited variation in absolute advantage across markets ($\sigma_\mu = 0.10$), which aligns with our observation of remarkably stable estimates for the parameters of the ability distribution across markets. Lastly, we estimate a substitution elasticity between sector goods within a market of $\rho = 2.72$, which is lower than the substitution elasticity within sector F and higher than the substitution elasticity across product markets, which is implicitly set to 1 in our case due to the Cobb-Douglas assumption, and is estimated at 1.5 in Gutiérrez (2023). This is consistent with gradually decreasing substitution elasticity as we move to upper utility nests.

4.4 Model Fit

We begin by evaluating how effectively the model replicates the distributions of key variables across local labor markets. Online Appendix Figure A.15 illustrates the distribution of normalized (log) sales, the (log) earnings gap between wage and self-employed workers, and the number of firms. The red bars represent the data, while the blue bars show the model's predicted distributions. Despite only targeting few key moments inestimation, the model's distributions closely match those observed in the data, demonstrating a good overall fit.

Model-Implied Reduced-Form Elasticities We now evaluate how well the model captures labor market power in the Peruvian economy. In Section 2.5, we provided evidence of significant labor market power across markets. However, those derived from the data are reduced-form inverse labor supply elasticity estimates, which incorporate competitors' employment responses and do not directly map to the structural elasticity (Berger, Herkenhoff and Mongey, 2022). To assess the model's fit, we replicate Table 2 using model simulations by applying the same shock to firm productivity and labor demand used in Section 2.5. This is challenging for several reasons, including the model's static nature, the wage homogeneity within markets, and the absence of electricity as a production input for firms.

[39]We find that in a version of the model where fixed costs are heterogeneous across markets and uncorrelated with firm productivity, we better replicate the lower correlation reported in Table 4. However, this adjustment comes at the cost of introducing additional assumptions—such as how fixed cost draws correlate with firm or worker productivity—and does not improve performance in other areas. In fact, the adjustment worsens outcomes in some respects and remains inconsequential beyond this specific aspect.

[40]This difference suggests a thinner productivity tail among domestic Peruvian producers compared to Chilean importers.

Table 5: Reduced-Form Inverse Supply Elasticity – Model Estimates

			Self-Empl. Rate	
			Low	High
	(1)	(2)	(3)	(4)
All Markets	0.364			
	(0.070)			
$HHI^{wn} \in (0, 0.18]$		0.242		
		(0.071)		
$HHI^{wn} \in (0.18, 0.25]$		0.319		
		(0.046)		
$HHI^{wn} \in (0, 0.25]$			0.214	0.284
			(0.033)	(0.134)
$HHI^{wn} \in (0.25, 1]$		0.560	0.621	0.479
		(0.160)	(0.177)	(0.294)

Notes. This table presents the estimates of the average inverse labor supply elasticity of treated firms, i.e., $\hat{\epsilon}_{iF,k} \equiv \Delta \ln W_{F,k} / \Delta \ln n_{iF,k}$, across all markets with more than one firms (Column 1) and within different market subsets (Columns 2 to 4) in the estimated model. These estimates are compared to the reduced-form markdown estimates provided in Table 2. The procedure to obtain these estimates are detailed in Online Appendix C.4. Low and high self-employment rates are defined as being below or above the average self-employment rate across local labor markets, respectively. Bootstrap standard errors are in parentheses. These are obtained by redrawing the iid shock associated with the assignment of τ values 1,000 times.

We overcome these challenges by following a three-step procedure, detailed in Online Appendix C.4. First, we identify the treated firms in the model by assigning an electricity wedge (τ) to each firm, based on its relationship with firm productivity, which we infer from the data. A firm is classified as treated if its τ exceeds the economywide median, consistent with the approach in Section 2.5. Second, we determine the magnitude of the productivity shock induced by electrification by estimating its effect on firm productivity in the data. Lastly, starting from the baseline model equilibrium, we simulate a 2.3% productivity shock to the treated firms, corresponding to the estimated average effect. The model's average reduced-form inverse labor supply elasticities are then calculated by taking the ratio of the (log) wage to (log) employment responses of treated firms in markets with more than one firms, consistent with the local average treatment effect (LATE) within-market estimates reported in Table 2.

Table 5 presents the model-implied reduced-form estimates of labor market power across markets, along with bootstrap standard errors, averaged across all markets as well as within different subsamples. The results align closely with those in Table 2. The model estimates an average inverse elasticity of 0.36 across local labor markets, compared to 0.42 in the data. The model effectively replicates the relationship between average inverse elasticity and market concentration, capturing its gradient with precision. Additionally, it reflects the mitigating effect of self-employment, showing that labor market power is highest in markets with high employer concentration and low self-employment rates. In these markets, the model predicts an inverse elasticity of 0.62, compared to 0.75 in the data. Conversely, labor market power is lower in concentrated markets with a higher-than-average self-employment share, though the model underestimates the difference here, overshooting the inverse elasticity estimate in this latter group of markets. We attribute this discrepancy to the model's inability to fully capture the broad variation in self-employment rates across markets, with a standard deviation of 0.09 in the model compared to 0.32 in the data. The difference-in-differences between highly concentrated and less concentrated markets with varying self-employment rates remains.

The estimated model also allows for comparison between reduced-form inverse elasticities and their structural counterparts, reported in Online Appendix Table A.11. The structural inverse elasticities are more subdued than the reduced-form estimates in Table 5, underscoring the importance of equilibrium responses from competitors. However, as discussed in Section 3.3, the bias in reduced-form estimates does not account for the heterogeneity observed across markets in terms of concentration or self-employment prevalence. Online Appendix Table A.11 supports this conclusion quantitatively. Additionally, Online Appendix Figure A.14 illustrates that no discernible pattern emerges when comparing structural and reduced-form labor market power against employer concentration and self-employment rates.

4.5 Robustness

Our structural approach to identifying labor market power relies on parametric assumptions. To validate the robustness of our findings, we explore several alternative parameterizations and their effect on labor market power and its variation across markets.

4.5.1 Heterogeneity in Roy Parameters

In our baseline specification, we assume a constant variance-covariance matrix and mean comparative advantage for workers' abilities across local labor markets. However, Online Appendix Figure A.12 shows that markets vary in their relative skill endowments, raising concerns that overlooking this heterogeneity may lead to biased estimates of labor market power. In particular, inspection of equations (15) and (16) reveals that variation in labor supply elasticity may be partially driven by these additional sources of heterogeneity, rather than self-employment shares alone.

To address this concern, we modify the model to allow for heterogeneity in both the variance-covariance matrix, $\boldsymbol{\Sigma_k}$, and the mean comparative advantage, $\hat{\mu}_k$. Specifically, we group markets into I clusters based on population quantiles. For each cluster $i = 1, \ldots, I$, we obtain the group-specific parameters $(\sigma_{F,i}, \sigma_{S,i}, \varrho_i, \hat{\mu}_i)$ as the within-group median. Markets in the model are then assigned to groups according to their population quantile. In the baseline model, we implictly set $I = 1$. In the robustness exercise, we set $I = 3$.

Online Appendix Table A.12 presents the estimated Roy parameters at baseline and under group heterogeneity, showing minimal variation across groups. Column 2 of Online Appendix Table A.13 reports the sensitivity of our labor market power estimates to this robustness check, demonstrating that the overall incidence of labor market power, as well as its relationship with market concentration and self-employment, remains largely unchanged. These findings support our decision to use a constant set of parameters for the baseline calibration.

4.5.2 Heterogeneity in Fixed Costs

Firm entry costs into the wage sector represent barriers to starting a formal firm, including regulatory procedures, high licensing fees, limited access to credit, and inadequate infrastructure. These barriers are likely heterogeneous across local labor markets. In our baseline calibration, we assumed a constant cost structure across markets, implying that entry—and therefore market concentration—was fully proportional to variable profits and productivity. This assumption may influence labor market power estimates.

To address this, we recalibrate the model allowing for an alternative parameterization of fixed costs, assuming each market draws f_k^e from a Weibull distribution with shape parameter f_κ and scale parameter f_λ. Column 3 of Online Appendix Table A.13 shows that our labor market power estimates remain robust under this assumption. Online Appendix Table A.14 demonstrates that this adjustment improves the model's ability to replicate the correlation between market concentration and earnings across local labor markets, particularly in highly concentrated markets where the baseline model tends to underestimate earnings.

However, this adjustment does not significantly affect the model's performance in capturing labor market power dynamics or its broader implications. In fact, while it addresses one specific issue, it introduces additional assumptions about how fixed costs correlate with productivity and it worsens the model's fit in other respects—namely the number of firms across markets—making it a less appealing solution overall.

4.5.3 Full Entry Game

In the baseline model, we adopt a simplified entry game to avoid the computational complexity of solving for exact equilibrium values of M_k, using an approximation where firms are treated as infinitesimally small at the entry stage. To ensure robustness, we recalibrate the model with a full entry game. Column 4 of Online Appendix Table A.13 demonstrates that even under this more complex entry framework, our labor market power estimates, as well as their co-movement with market concentration and self-employment, remain unaffected. This is reassuring, as we argued in Section 4.1 that the entry assumption primarily influences equilibrium through the number of firms and market concentration, which we target in calibration, without affecting other core outcomes of the model.

5 Counterfactual Policy Analyses

Armed with the estimated model, we conduct two sets of counterfactual experiments to address our key research questions. First, we quantitatively assess the role of labor market power in shaping labor market outcomes in Peru. Second, we simulate three industrial policies aimed at promoting industrialization and increasing wage employment by targeting firms or workers. We measure their aggregate and distributional impacts and investigate how labor market power affects the success of these policies.

5.1 Impact of Labor Market Power

To investigate the impact of labor market power on labor market outcomes, we introduce a conduct parameter, $\iota \equiv \frac{dN_{F,k}}{dn_{iF,k}} = \{0, 1\}$, which captures the firm's perceived effect of its labor demand on aggregate market variables. When $\iota = 1$, firms are strategic and fully internalize the impact of their labor demand on aggregate labor demand and wages, as in our baseline model. Conversely, when $\iota = 0$, firms behave as wage-takers, irrespective of market concentration.

Under this generalization of market conduct, the firm i's markdown in market k becomes:

$$\psi_{iF,k} = 1 + \iota \frac{s_{iF,n}}{\epsilon(\hat{W}_k)}, \tag{21}$$

which highlights that the markdown equals one whenever firms act as wage takers.[41] We quantify the role of labor market power in the economy by comparing the baseline economy with one where $\iota = 0$, holding other things equal.

Figure 3 presents some key distributions of interest, while Online Appendix Table A.16 shows the average outcomes in the two economies. Panel (a) of Figure 3 compares the distribution of wage employment across markets, showing that labor market power significantly limits wage employment in Peru. With perfectly competitive labor markets, the share of wage employment would increase by over ten percentage points, from from 66 to 77%.

Panel (b) shows that despite increased competition, concentration persists and even rises in the absence of labor market power due to market share reallocation. Without labor market power, the most productive firms gain market share, leading to higher concentration.[42]

Panel (c) shows that labor market power depresses average wages, while panel (d) reveals that it narrows the earnings gap between wage workers and the self-employed. Without labor market power, average earnings would rise by 31% in the wage employment sector, and by 27% in the self-employment sector.

Drawing on insights from Section 3.3, we can decompose the changes in average earnings in both sectors into their key components.[43] The increase in average wages in the no-labor-market-power economy is entirely driven by higher earnings per efficiency unit, with the elimination

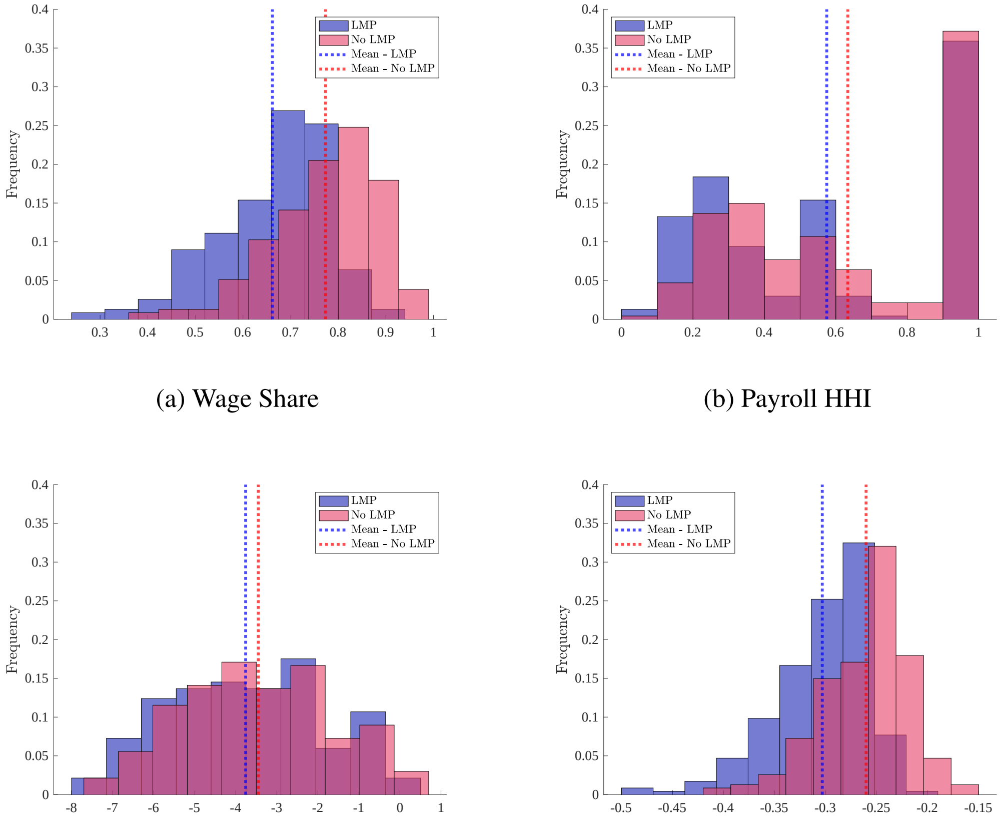
[41] The wage markdown of firm i in market k can be written as:

$$\psi_{iF,k} = 1 + \frac{d \ln W_{F,k}}{d \ln n_{iF,k}} = 1 + \frac{d \ln W_{F,k}}{d \ln N_{F,k}} \cdot \frac{dN_{F,k}}{dn_{iF,k}} \cdot \frac{n_{iF,k}}{N_{F,k}} = 1 + \iota \frac{s_{iF,n}}{\epsilon(\hat{W}_k)}.$$

[42] See Eslava, Haltiwanger and Urdaneta (2024) for empirical evidence of this channel.

[43] These results are detailed in Online Appendix Table A.16.

Figure 3: Effects of Labor Market Power



(a) Wage Share

(b) Payroll HHI

(c) Average Wages (Log)

(d) Earnings Gap (Log)

Notes. The four panels show the distribution of key labor market outcomes across markets in the baseline (blue) and in the counterfactual economy (red) with no labor market power. Online Appendix Table A.16 complements these figures by reporting the average of selected outcomes across markets in the two economies together with the difference between the two.

of markdowns accounting for a 35% wage increase. This is partially offset by a 4% reduction in labor revenue productivity (MRPL). As market share shifts toward more productive, high-markup firms, output prices fall, accounting for most of the decline in MRPL. The selection channel has no impact, as expected from our parameter estimates.

In the self-employment sector, average earnings also rise, and about two-thirds of the increase is driven by higher unit earnings, fully attributable to higher output prices. Unlike in the wage sector, the selection channel plays a crucial role here. In the counterfactual economy, the average ability of self-employed workers is 11% higher, as more workers transition to wage employment, leaving the remaining self-employed workers increasingly positively selected.

Overall, the results in this section demonstrate that labor market power has a strong hold on the Peruvian economy. It contributes to the scarcity of wage jobs, lowers firm size, and reduces wages and self-employment earnings through markdowns, selection, and revenue productivity effects. Our findings also show that worker self-selection is a key factor through which labor market power in wage employment reduces earnings in the self-employment sector, thereby influencing the earnings gap between wage workers and the self-employed.

5.2 Industrial Policy

Despite long-standing efforts to increase wage employment as a means of promoting inclusive growth, policy interventions have often had limited impact (Bandiera et al., 2022). This section examines the extent to which the interaction between labor market power and self-employment contributes to this outcome. We simulate three industrial policy interventions—targeting firm productivity, worker productivity, and entry costs—and use our model to evaluate how labor market power affects their overall effectiveness, both qualitatively and quantitatively.

5.2.1 Firm Productivity

Policy efforts to boost firm productivity have often focused on market integration, primarily through infrastructure improvements (Fiorini, Sanfilippo and Sundaram, 2021). The goal is to expand market access, enhance productivity, and reduce both information frictions and shipping costs for inputs and outputs. To assess such interventions, we examine a road infrastructure project in Peru. Between 2003 and 2010, the country added over 5,000 kilometers of new roads, expanding the network by more than 10%. Volpe Martincus, Carballo and Cusolito (2017) evaluate this intervention and find that firm exports increased by an average of 3.7% as a result. Building on this evidence, we calibrate a shock in our model that shifts expected market-level productivity (T_k) to achieve a comparable increase in firm sales across markets.[44]

Although the shock is applied uniformly across markets, its effects vary significantly. Panel (a) of Figure 4 shows the estimated average impact across markets, segmented by quintiles of the baseline wage markdown distribution.[45] Wage employment shares and wages rise on average across all market groups, as intended by the policy. However, the rise in productivity does not fully translate into higher wages. The incomplete pass-through stems from an increase in wage markdowns, resulting from the policy's effects on its two key determinants: employer concentration and labor supply elasticity. As firm productivity rises, market entry increases, which reduces concentration. Yet, the labor supply elasticity also decreases as more workers opt for wage employment. The latter effect dominates in most markets, as shown by Online Appendix Figure A.16, leading to higher wage markdowns.[46]

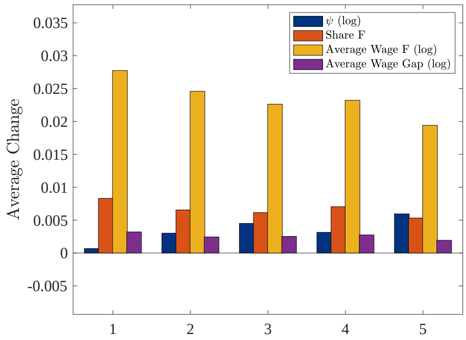
Second, despite higher markdowns and stronger positive selection into self-employment,

[44] We estimate that T_k must increase by 8% to reach the targeted sales growth.

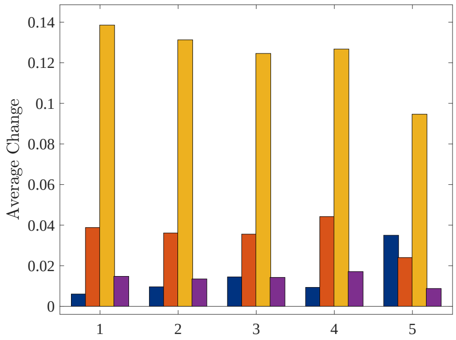
[45] Online Appendix Table A.17 provides average changes in selected outcomes across all markets.

[46] We also compare the effects of this policy in general equilibrium (GE) versus a partial equilibrium (PE) setting, where aggregate income is fixed. The results shown in Online Appendix Table A.15 reveal significant differences: in the PE model, a productivity shock leads to declining unit wages due to fixed demand and price reductions, causing lower wages in both sectors as workers reallocate. Conversely, the GE model shows that falling prices boost aggregate demand, increasing the marginal revenue product of labor and raising wages across all sectors. This underscores the importance of GE structure in accurately capturing wage responses, as the GE model significantly mitigates the negative wage impacts observed in the PE scenario.

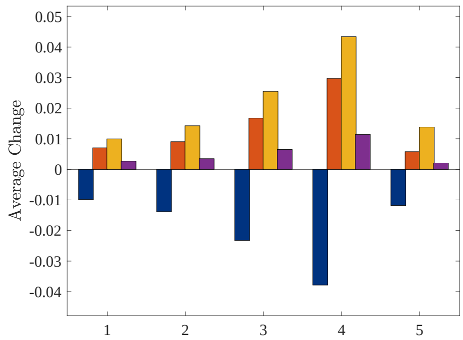
Figure 4: Effects of Policy Shocks Across Markets



(a) Firm Productivity



(b) Worker Skills



(c) Fixed Cost

Notes. The three panels illustrate the estimated change in wage markdown, wage employment share, average wage, and average earning gap between wage workers and self-employed workers 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. Online Appendix Figure A.16 complements these figures by showing the change in the wage markdown together with its determinants, i.e. employer concentration and wage work supply elasticity.

the earnings gap between wage and self-employed workers increases moderately. This occurs because the rising productivity of formal firms, combined with general equilibrium effects, more than offsets the opposing effects of markdowns and selection.

Third, and most importantly, the figure demonstrates that labor market power and its determinants strongly influence the policy's impact. Increasing firm productivity is significantly more effective in markets with lower baseline wage markdowns. This is true despite the higher baseline levels of wage employment shares and wages in these markets, as discussed in Section 5.1. The average increase in wage employment share and wages is 57% and 43% higher, respectively, in markets in the bottom quintile of the baseline wage markdown distribution compared to those in the top quintile.[47] Most notably, rising markdowns account for 98% of the variation in the policy's impact on wages and 85% of the variation in its effect on the wage

[47] Wages, for instance, increase by 1.94% on average in the least competitive labor markets and by 2.77% in the most competitive ones, so $(2.77 - 1.94)/1.94 = 0.43$.

employment share across markets.[48]

5.2.2 Worker Skills

Next, we examine policies aimed at boosting the supply of wage labor by enhancing worker skills through targeted training programs. These initiatives operate on the belief that unemployment stems from a lack of specific technical skills, which can be addressed through short-term training (McKenzie, 2017). Several such programs have been implemented across Latin America, including Peru's Job Youth Training Program, known as *Projoven*. Operating from 1996 to 2010, the program aimed to equip young people from low-income backgrounds with training and labor market experience aligned with the needs of the productive sector, catering directly to employer demands. An experimental evaluation of Projoven by Díaz and Rosas-Shady (2016) found that, two years after completing the program, participants had a 3.6 percentage point higher probability of securing wage employment compared to the control group, although the result was not statistically significant.[49] To simulate a similar training program in our model, we introduce a shock to workers' mean comparative advantage in wage employment, $\hat{\mu}$.[50]

Panel (b) of Figure 4 shows that, by improving worker skills, the program raises productivity in the formal sector, similar to a productivity shock to firms. This reduces concentration and increases wage employment. Additionally, as worker skills improve, the mean comparative advantage in wage employment increases, making self-employment less attractive and decreasing the supply elasticity of wage work, as shown in Online Appendix Figure A.16. This reduction in elasticity outweighs the decline in concentration, resulting in higher wage markdowns. Despite this, wages still increase as skills improve. Notably, the model closely replicates the 13.4% rise in monthly earnings reported by Díaz and Rosas-Shady (2016).

Labor market power and its changes play a crucial role in shaping the policy's impact. The average effects on wage employment and wages are 62% and 47% higher, respectively, in the most competitive labor markets compared to the least competitive ones. As with policies that boost firm productivity, rising markdowns account for 98% of the variation in the policy's effect on average wages and 85% of the variation in its impact on the wage employment share.

5.2.3 Firm Entry Cost

Policies to reduce entry costs typically involve government programs that simplify entry regulations. A prime example is the Mexican Rapid Business Opening System (SARE), which aimed to streamline local business registration procedures across various municipalities starting in

[48]These figures are derived from the R^2 of a simple regression implemented on simulated data, where the local labor market is the unit of observation. The dependent variable is the change in average wage or wage employment share following the policy shock, and the independent variable is the change in the wage markdown.

[49]However, significant positive effects were found in the likelihood of obtaining formal employment, such as jobs with health insurance and pensions.

[50]In order to increase average wage employment by 3.6 percentage points, $\hat{\mu}$ must rise from -0.12 to 0.04 across all markets.

May 2002. Both Kaplan, Piedra and Seira (2011) and Bruhn (2011) evaluate the impact of this reform on several economic outcomes at the municipality level.[51] For our policy experiment, we target the 2.2% increase in the fraction of wage earners documented in Bruhn (2011).[52]

Panel (c) of Figure 4 shows the policy impact across markets. The policy encourages firm entry and reduces concentration. Competition intensifies in the output and labor market, leading to lower markups and prices.[53] Wage employment increases, but the effect is modest compared to the large reduction in concentration. This is because the drop in concentration stems from the entry of relatively unproductive firms, which have little impact on aggregate labor demand. As a result, the policy reduces concentration more than it affects labor supply elasticity, and wage markdowns decrease, as shown in Online Appendix Figure A.16.

The policy's impact is highly heterogeneous across markets. However, unlike the firm productivity shock, reducing entry costs is least effective in the most competitive labor markets. It is also relatively ineffective in the most concentrated and least competitive markets. In markets with moderately high markdowns, reducing entry costs is more impactful and effective because concentration is on the margin more responsive to the negative cost shift.

Yet, as in the previous two cases, changes in labor market power are crucial to understanding the effects of reducing fixed costs. Changes in markdowns explain 99% of the variation in the policy's impact on wages and 88% of the variation in its effect on the wage employment share.

5.2.4 Discussion

These counterfactual exercises show that our framework provides a valuable perspective for understanding the impact of industrialization policies. The effectiveness of these policies is closely tied to labor market power and its key drivers—employer concentration and labor supply elasticity. While these policies help create more wage jobs, they also make self-employment less attractive, reducing labor supply elasticity. In some cases, this means that pro-competitive policies can have the unintended effect of making the labor market less competitive. For productivity-enhancing policies, the drop in elasticity outweighs the reduction in concentration, leading to higher wage markdowns and weakening the policy's overall impact. In contrast, policies that lower entry costs reduce both concentration and markdowns, making them more effective in markets with moderate labor market power. In all cases, changes in markdowns largely explain the variation in policy outcomes across markets. These insights are crucial for researchers and policymakers seeking to better understand labor market dynamics and develop effective strategies for industrialization and inclusive growth.

[51] The main difference between the two studies is that Bruhn (2011) uses household data from labor market surveys, whereas Kaplan, Piedra and Seira (2011) uses social security data.

[52] We find that a 40% reduction in fixed costs across local labor markets is required to achieve an average 2.2% increase in wage employment in our model.

[53] Remarkably, the model replicates the 1% decrease in prices found by Bruhn (2011). Specifically, this refers to the average change across markets in the price index in the wage employment sector, $P_{F,k}$, as reported in Online Appendix Table A.17.

6 Conclusions

Addressing the scarcity of good jobs in poor countries remains a major challenge for both policymakers and researchers. This paper emphasizes the crucial role that labor market power, and its interaction with self-employment, plays in shaping labor market outcomes. Drawing on new evidence from Peru, a general equilibrium model, and counterfactual policy experiments, we argue that understanding this interaction is essential for designing effective interventions.

Our findings reveal that self-employment plays a dual role in the presence of labor market power. On one hand, it acts as a safety net, providing an alternative when wage employment is scarce. On the other hand, it undermines industrial policies aimed at promoting wage employment. The variable elasticity of labor supply is key to these dynamics, and this paper makes a significant contribution by using worker-level data to structurally identify this elasticity. This channel proves quantitatively critical for understanding the broader impact of labor market power on the economy and policy effectiveness.

Moreover, our results challenge the traditional view that the self-employment sector is a perfectly elastic reservoir of labor for industrialization (Lewis, 1954; Rauch, 1991). Instead, we argue that labor market power can obstruct the efficient transfer of workers into the industrial sector, resulting in an over-reliance on self-employment, which may slow the development process. This suggests that industrial development strategies must account for the interaction between labor market power and self-employment to be effective.

The implications of these findings extend beyond the context of low-income countries. The rise of the digital economy in wealthier nations has transformed the nature of firms and work, increasing self-employment and flexible work arrangements while reducing traditional employment. At the same time, there is growing evidence of significant labor market power held by firms in rich countries. Understanding how this shift interacts with the labor market power of traditional employers presents a promising avenue for future research, with important implications for labor markets worldwide.

Lastly, while this paper focuses on manufacturing, which has historically driven economic development and been the main target of industrial policy in low- and middle-income countries, it is essential to expand the analysis to other sectors, such as agriculture and services, which are critical to the economies of developing nations. Incorporating these broader sectors into our framework represents a natural and necessary extension of this research.

References

Abeberese, Ama Baafra, Charles Godfred Ackah, and Patrick Opoku Asuming. 2019. "Productivity Losses and Firm Responses to Electricity Shortages: Evidence from Ghana." *The World Bank Economic Review*, 35(1): 1–18.

Adão, Rodrigo. 2016. "Worker heterogeneity, wage inequality, and international trade: Theory and evidence from brazil." Mimeo.

Alfaro, Laura, Marcela Eslava, Martí Mestieri, and Luis Felipe Saenz. 2023. "Jobless Industrialization." Mimeo.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P. Vasquez. 2021. "The Effects of Multinationals on Workers: Evidence from Costa Rican Microdata." Mimeo.

Alfonsi, Livia, Oriana Bandiera, Vittorio Bassi, Robin Burgess, Imran Rasul, Munshi Sulaiman, and Anna Vitali. 2020. "Tackling youth unemployment: Evidence from a labor market experiment in Uganda." *Econometrica*, 88(6): 2369–2414.

Amodio, Francesco, and Nicolas De Roux. 2022. "Measuring Labor Market Power in Developing Countries: Evidence from Colombian Plants." *Journal of Labor Economics*. Forthcoming.

Amodio, Francesco, Emanuele Brancati, Peter Brummond, Nicolas de Roux, and Michele Di Maio. 2024. "Global Labor Market Power." CEPR Discussion Papers 18828.

Amodio, Francesco, Francisco Alvarez-Cuadrado, and Markus Poschke. 2020. "Selection and Absolute Advantage in Farming and Entrepreneurship." CEPR Discussion Papers 14269.

Armangué-Jubert, Tristany, Nezih Guner, and Alessandro Ruggieri. 2024. "Labor Market Power and Development." *American Economic Review: Insights*. Forthcoming.

Atkeson, Andrew, and Ariel Burstein. 2008. "Pricing to Market, Trade Costs, and International Relative Prices." *The American Economic Review*, 98(5): 1998–2031.

Azar, José, Ioana Marinescu, and Marshall Steinbaum. 2022. "Labor Market Concentration." *Journal of Human Resources*, 57(S): 167–199.

Bandiera, Oriana, Ahmed Elsayed, Andrea Smurra, and Céline Zipfel. 2022. "Young Adults and Labor Markets in Africa." *Journal of Economic Perspectives*, 36(1): 81–100.

Bassier, Ihsaan. 2023. "Firms and inequality when unemployment is high." *Journal of Development Economics*, 161: 103029.

Bassier, Ihsaan, Arindrajit Dube, and Suresh Naidu. 2022. "Monopsony in Movers: The Elasticity of Labor Supply to Firm Wage Policies." *Journal of Human Resources*, 57(S): 50–86.

Bau, Natalie, and Adrien Matray. 2023. "Misallocation and Capital Market Integration: Evidence From India." *Econometrica*, 91(1): 67–106.

Benmelech, Efraim, Nittai K. Bergman, and Hyunseob Kim. 2022. "Strong Employers and Weak Employees: How Does Employer Concentration Affect Wages?" *Journal of Human Resources*, 57(S): 200–250.

Berger, David, Kyle Herkenhoff, and Simon Mongey. 2022. "Labor Market Power." *American Economic Review*, 112(4): 1147–93.

Blattman, Christopher, and Stefan Dercon. 2018. “The Impacts of Industrial and Entrepreneurial Work on Income and Health: Experimental Evidence from Ethiopia.” *American Economic Journal: Applied Economics*, 10(3): 1–38.

Breza, Emily, Supreet Kaur, and Yogita Shamdasani. 2021. “Labor Rationing.” *American Economic Review*, 111(10): 3184–3224.

Brooks, Wyatt J., Joseph P. Kaboski, Yao Amber Li, and Wei Qian. 2021. “Exploitation of labor? Classical monopsony power and labor’s share.” *Journal of Development Economics*, 150: 102627.

Bruhn, Miriam. 2011. “License to Sell: The Effect of Business Registration Reform on Entrepreneurial Activity in Mexico.” *The Review of Economics and Statistics*, 93(1): 382–386.

Dasso, Rosamaría, and Fernando Fernandez. 2015. “The effects of electrification on employment in rural Peru.” *IZA Journal of Labor & Development*, 4(6).

Dasso, Rosamaría, Fernando Fernandez, and Hugo R. Ñopo. 2015. “Electrification and Educational Outcomes in Rural Peru.” Institute of Labor Economics (IZA) IZA Discussion Papers 8928.

Deb, Shubhdeep, Jan Eeckhout, Aseem Patel, and Lawrence Warren. 2022. “What drives Wage Stagnation: Monopsony or Monopoly?” *Journal of the European Economic Association*. jvac060.

Díaz, Juan José, and David Rosas-Shady. 2016. “Impact Evaluation of the Job Youth Training Program Projoven.” Inter-American Development Bank Working Paper IDB-WP-693.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2019. “Margins of labor market adjustment to trade.” *Journal of International Economics*, 117(C): 125–142.

Dix-Carneiro, Rafael, Pinelopi Koujianou Goldberg, Costas Meghir, and Gabriel Ulyssea. 2024. “Trade and Domestic Distortions: The Case of Informality.” Cowles Foundation for Research in Economics, Yale University Cowles Foundation Discussion Papers 2384.

Djankov, Simeon, Rafael La Porta, Florencio Lopez-de Silanes, and Andrei Shleifer. 2002. “The regulation of entry.” *The Quarterly Journal of Economics*, 117(1): 1–37.

Donovan, Kevin, Will Jianyu Lu, and Todd Schoellman. 2023. “Labor Market Dynamics and Development.” *The Quarterly Journal of Economics*, 138(4): 2287–2325.

Duranton, Gilles, and Henry G. Overman. 2005. “Testing for Localization Using Micro-Geographic Data.” *The Review of Economic Studies*, 72(4): 1077–1106.

Eaton, Jonathan, Samuel S. Kortum, and Sebastian Sotelo. 2012. “International Trade: Linking Micro and Macro.” NBER Working Papers 17864.

Ellison, Glenn, and Edward L. Glaeser. 1997. “Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach.” *Journal of Political Economy*, 105(5): 889–927.

Eslava, Marcela, John Haltiwanger, and Nicolas Urdaneta. 2024. “The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand, and Wedges.” *The Review of Economic Studies*, 91(1): 259–300.

Felix, Mayara. 2022. “Trade, Labor Market Concentration, and Wages.” Mimeo.

Fields, Gary. 2012. *Working Hard, Working Poor: A Global Journey.* Oxford University Press.

Fiorini, Matteo, Marco Sanfilippo, and Asha Sundaram. 2021. “Trade liberalization, roads and firm productivity." *Journal of Development Economics*, 153(C).

French, Eric, and Christopher Taber. 2011. "Identification of models of the labor market." *Handbook of labor economics*, 4: 537–617.

Gaubert, Cecile, and Oleg Itskhoki. 2021. "Granular comparative advantage." *Journal of Political Economy*, 129(3): 871–939.

Gollin, Douglas. 2008. "Nobody's business but my own: Self-employment and small enterprise in economic development." *Journal of Monetary Economics*, 55(2): 219 – 233.

Gutiérrez, Agustín. 2023. "Labor Market Power and the Pro-competitive Gains from Trade." Mimeo.

Harris, John R., and Michael P. Todaro. 1970. "Migration, Unemployment and Development: A Two-Sector Analysis." *The American Economic Review*, 60(1): 126–142.

Heckman, James J, and Guilherme Sedlacek. 1985. "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market." *Journal of Political Economy*, 93(6): 1077–1125.

Hjort, Jonas, Hannes Malmberg, and Todd Schoellman. 2022. "The Missing Middle Managers: Labor Costs, Firm Structure, and Development." NBER Working Papers 30592.

Hsieh, Chang-Tai, and Benjamin A Olken. 2014. "The missing "missing middle"." *Journal of Economic Perspectives*, 28(3): 89–108.

Huang, Hanwei, Kalina Manova, Oscar Perello, and Frank Pisch. 2024. "Firm Heterogeneity and Imperfect Competition in Global Production Networks." Mimeo.

Huneeus, Federico, and Richard Rogerson. 2023. "Heterogeneous Paths of Industrialization." *The Review of Economic Studies*, 91(3): 1746–1774.

International Labour Organization. 2020. "ILO modelled estimates database."

Juhász, Réka, Nathan J. Lane, and Dani Rodrik. 2023. "The New Economics of Industrial Policy." NBER Working Papers 31538.

Kaboski, Joseph P., and Robert M. Townsend. 2011. "A structural evaluation of a large-scale quasi-experimental microfinance initiative." *Econometrica*, 79(5): 1357–1406.

Kahn, Matthew E., and Joseph Tracy. 2024. "Monopsony in spatial equilibrium." *Regional Science and Urban Economics*, 104(C).

Kaplan, David S., Eduardo Piedra, and Enrique Seira. 2011. "Entry regulation and business start-ups: Evidence from Mexico." *Journal of Public Economics*, 95(11): 1501–1515.

La Porta, Rafael, and Andrei Shleifer. 2014. "Informality and Development." *Journal of Economic Perspectives*, 28(3): 109–26.

Lewis, W. Arthur. 1954. "Economic Development with Unlimited Supplies of Labour." *The Manchester School*, 22(2): 139–191.

MacKenzie, Gaelan. 2021. "Trade and Market Power in Product and Labor Markets." Bank of Canada Staff Working Papers 21-17.

Maloney, William F. 1999. "Does informality imply segmentation in urban labor markets? Evidence from sectoral transitions in Mexico." *The World Bank Economic Review*, 13(2): 275–302.

Manning, Alan. 2003. *Monopsony in Motion: Imperfect Competition in Labor Markets.* Princeton University Press.

McKenzie, David. 2017. "How Effective Are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence." *The World Bank Research Observer*, 32(2): 127–154.

McMillan, Margaret, and Albert Zeufack. 2022. "Labor Productivity Growth and Industrialization in Africa." *Journal of Economic Perspectives*, 36(1): 3–32.

Muralidharan, Karthik, Paul Niehaus, and Sandip Sukhtankar. 2023. "General Equilibrium Effects of (Improving) Public Employment Programs: Experimental Evidence From India." *Econometrica*, 91(4): 1261–1295.

Méndez, Esteban, and Diana Van Patten. 2022. "Multinationals, Monopsony, and Local Development: Evidence From the United Fruit Company." *Econometrica*, 90(6): 2685–2721.

Pham, Hoang. 2023. "Trade reform, oligopsony, and labor market distortion: Theory and evidence." *Journal of International Economics*, 144: 103787.

Piselli, Roberto. 2013. "Patrones de transporte en Lima Metropolitana: adónde, cuánto y por qué viajan los limenos,." *Revista Argumentos*, 25–28.

Ponczek, Vladimir, and Gabriel Ulyssea. 2021. "Enforcement of Labour Regulation and the Labour Market Effects of Trade: Evidence from Brazil." *The Economic Journal*. ueab052.

Poschke, Markus. 2022. "Wage employment, unemployment and self-employment across countries." Mimeo.

Rauch, James. 1991. "Modelling the informal sector formally." *Journal of Development Economics*, 35(1): 33–47.

Roy, Andrew Donald. 1951. "Some thoughts on the distribution of earnings." *Oxford economic papers*, 3(2): 135–146.

Rud, Juan Pablo, and Ija Trapeznikova. 2021. "Job creation and wages in least developed countries: Evidence from sub-saharan africa." *The Economic Journal*, 131(635): 1331–1364.

Ulyssea, Gabriel. 2018. "Firms, Informality, and Development: Theory and Evidence from Brazil." *American Economic Review*, 108(8): 2015–47.

Ulyssea, Gabriel. 2020. "Informality: Causes and Consequences for Development." *Annual Review of Economics*, 12(1): 525–546.

UN General Assembly. 2015. "Transforming our world: the 2030 Agenda for Sustainable Development." *United Nations: New York, NY, USA*.

Volpe Martincus, Christian, Jerónimo Carballo, and Ana Cusolito. 2017. "Roads, exports and employment: Evidence from a developing country." *Journal of Development Economics*, 125(C): 21–39.

Wong, Horng Chern. 2023. "Understanding High-Wage Firms." Mimeo.

Yeh, Chen, Claudia Macaluso, and Brad Hershbein. 2022. "Monopsony in the US Labor Market." *American Economic Review*, 112(7): 2099–2138.

Online Appendix

Labor Market Power, Self-Employment, and Development

Francesco Amodio, Pamela Medina, Monica Morlacco*

This Appendix is organized as follows. Section A provides all the additional tables and figures referenced in the text. Section B integrates Section 3 by showing additional theoretical results. Section C complements Section 4 and provides further details on the estimation procedures.

*Amodio: francesco.amodio@mcgill.ca, McGill University and CIREQ, 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: Employer Concentration Across Local Labor Markets
Alternative Samples and Definitions

	Full Sample		*Merged Sample*		*Broadened LLMs*	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Number of Firms	6.39	10.37	7.25	11.19	10.24	22.83
Wage-bill HHI	0.65	0.33	0.61	0.34	0.68	0.35
Wage-bill HHI (Payroll Weighted)	0.37	0.03	0.37	0.03	0.26	0.03
Wage-bill HHI (Employment Weighted)	0.34	0.03	0.34	0.03	0.34	0.03
Employment HHI	0.63	0.35	0.59	0.35	0.66	0.36
Employment HHI (Payroll Weighted)	0.33	0.03	0.33	0.03	0.21	0.02
Employment HHI (Employment Weighted)	0.31	0.02	0.31	0.02	0.21	0.02
Percent of LLMs with 1 firm	38.78	2.27	38.78	2.29	44.65	2.58
Payroll Share of LLMs with 1 firm	7.94	1.79	7.96	1.81	6.65	1.72
Employment Share of LLMs with 1 firm	7.80	1.23	7.81	1.25	6.38	1.10
Number of Local Labor Markets	280		228		179	
Number of Locations	61		48		23	
Industries	23		22		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, the subset merged with worker-level data from ENAHO, and the full sample where local labor markets are more broadly defined as 2-digit industries within Peruvian departments. In the first two groups, local labor markets are defined by 2-digit industries within locations, with locations corresponding to Peruvian provinces or commuting zones.

Table A.2: 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 A.3: 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	0.25	0.11	0.25	0.12	0.32	0.37	0.12
Mining	0.01	0.02	0.03	0.02	0.00	0.01	0.00
Manufacturing	0.11	0.13	0.08	0.31	0.09	0.07	0.28
Utilities	0.00	0.01	0.01	0.02	0.00	0.00	0.00
Construction	0.05	0.07	0.14	0.12	0.04	0.08	0.03
Retails	0.15	0.05	0.04	0.05	0.25	0.14	0.20
Transportation	0.06	0.04	0.04	0.02	0.09	0.10	0.11
Other Services	0.36	0.57	0.42	0.35	0.21	0.24	0.25
Observations	234174	88669	1736	113	118143	1746	162

Notes. This table shows the distribution of employment across sectors in ENAHO amongst all workers, wage workers, and self-employed workers, averaging across all years from 2004 to 2011. Using the 2007-2011 panel, it also shows the distribution of wage work (and self-employment) among workers who were self-employed (wage workers) in the previous year, both across sectors and in manufacturing.

Table A.4: 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.5: 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 A.6: 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 A.7: 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.

Table A.8: Estimates of Labor Market Power – First Stage Regression Results

$HHI^{wn} \in$	$(0, 1]$	$(0, 0.18]$	$(0.18, 0.25]$	$(0.25, 1]$	$(0, 0.25]$	$(0, 0.25]$	$(0.25, 1]$	$(0.25, 1]$
Self-employment Rate	All	All	All	All	Low	High	Low	High
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Log of Employment							
$PER_{gt} \times EC_{i(j,g)}$	0.005***							
	(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)				
$\times \mathbb{I}\{HHI^{wn} \in (0, 0.25]\} \times \mathbb{I}\{$Low SE$\}$					-0.010***	0.009***	0.003***	0.002**
					(0.001)	(0.000)	(0.000)	(0.001)
$\times \mathbb{I}\{HHI^{wn} \in (0, 0.25]\} \times \mathbb{I}\{$High SE$\}$					0.012***	-0.010***	0.003***	0.002**
					(0.001)	(0.001)	(0.000)	(0.001)
$\times \mathbb{I}\{HHI^{wn} \in (0.25, 1]\} \times \mathbb{I}\{$Low SE$\}$					0.001	0.003	-0.009***	0.010***
					(0.001)	(0.003)	(0.002)	(0.003)
$\times \mathbb{I}\{HHI^{wn} \in (0.25, 1]\} \times \mathbb{I}\{$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.9: 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	312.60	848.74	282.65	1065.56
		1545.11	446.01	1211.54
		14292.13		
Observations	4954	4954	3257	1697

Notes. * p-value$<$ 0.1; ** p-value$<$0.05; *** p-value$<$0.01. The unit of observation is a medium to a large firm in EEA. The table reports 2SLS estimates of the firm-level inverse elasticity of supply of wage work as captured by β in equation (1). The instrumental variable is the interaction of the cumulative number of PER projects completed in each location g up to year t (PER_{gt}) and a dummy equal to one for firms with higher than median constraints to accessing electricity at baseline ($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.10: Estimates of Labor Market Power – First Stage Regression Results
Robustness to District $\times$ Industry $\times$ Year Fixed Effects

$HHI^{wn} \in$ 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				
$PER_{gt} \times EC_{i(j,g)}$	0.006***							
	(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.9. Columns 2 to 4 report the first-stage results from the three first-stage regressions associated with column 2 of Online Appendix Table A.9. Columns 5 to 8 report those from the four first-stage regressions associated with column 3 and 4 of Online Appendix Table A.9. 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.11: 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.12: 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	-0.12	-0.37	-0.42
Panel II. Group Heterogeneity						
Group 1	0.91	0.93	0.87	-0.27	-0.5	-0.53
Group 2	0.83	0.91	0.96	-0.17	-0.17	-0.21
Group 3	0.69	0.91	0.88	-0.27	-0.37	-0.39

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 C.1.2, with $\ln \frac{\text{Earnings}_F}{\text{Earnings}_S}$ as our baseline.

Table A.13: Labor Market Power – Robustness

	(1) Baseline	(2) Roy Estimates	(3) Fixed Costs	(4) Entry
$\bar{\psi}_k$	1.45	1.42	1.46	1.46
— High HHI	1.71	1.67	1.76	1.72
— High self-employment	1.40	1.42	1.37	1.40

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.14: Targeted Moments and Model Fit

Moment	Baseline	Fixed Costs	Entry
Panel I. Distribution Moments			
Log Number of Firms			
Mean	0.97	2.58	1.01
Standard Deviation	0.95	1.79	0.85
Log of Sales			
Ratio p75/p25	2.94	2.90	2.95
Ratio p90/p10	5.29	5.19	5.27
CR_1, Mean	0.66	0.68	0.68
CR_1, Standard Deviation	0.29	0.26	0.26
CR_4, Mean	0.94	0.93	0.96
CR_4, Standard Deviation	0.11	0.10	0.09
Employment HHI			
Mean, Unweighted	0.57	0.52	0.56
Standard Deviation	0.34	0.32	0.31
Mean, Weighted	0.30	0.41	0.31
Percent of Markets with 1 firm	0.36	0.21	0.29
Wage-bill Share of			
Markets with 1 firm	0.07	0.08	0.01
Markets with $<$10 firms	0.89	0.41	0.91
Markets with $<$50 firms	1.00	0.62	1.00
Share of Wage Employment			
Mean	0.66	0.69	0.67
Standard Deviation	0.12	0.11	0.11
Log of Earnings$_F$/Earnings$_S$			
Mean	0.41	0.57	0.47
Standard Deviation	0.58	0.55	0.55
Log of Schooling$_F$/Schooling$_S$ (Ability$_F$/Ability$_S$)			
Ratio p75/p25	1.42	1.29	1.33
Ratio p90/p10	1.18	1.14	1.16
Panel II. Regression Coefficients			
% Wage Employment on (Log) HHI^{wb}			
Point Estimate	-0.04	-0.03	-0.04
Standard Error	0.01	0.01	0.01
(Log) Earnings$_S$ on (Log) HHI^{wb}			
Point Estimate	-1.17	-0.13	-1.60
Standard Error	0.12	0.14	0.12
(Log) Earnings$_F$ on (Log) HHI^{wb}			
Point Estimate	-1.36	-0.26	-1.79
Standard Error	0.13	0.14	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.15: 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.16: 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 Work Supply Elasticity $\epsilon(\hat{W}_k)$	0.24	-0.18	-0.42

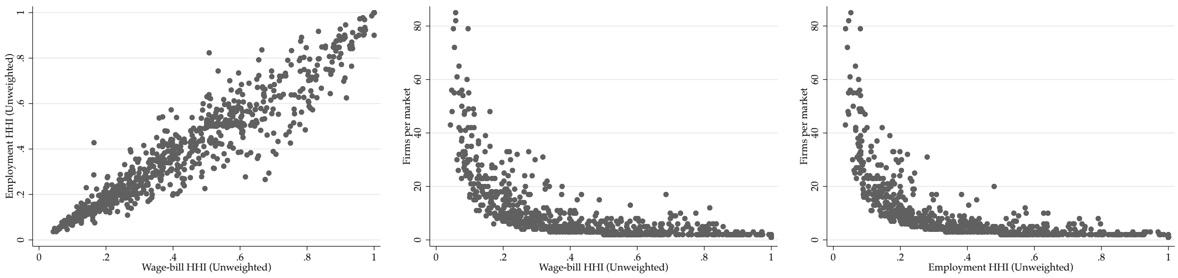
Notes. This table reports the average of selected outcomes across markets in the baseline economy ($\bar{Y}_{\iota=1}$) and in the counterfactual economy with no labor market power ($\bar{Y}_{\iota=0}$) together with the difference between the two.

Table A.17: 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 Work Supply Elasticity $\epsilon(\hat{W}_k)$	-1.91	-3.69	-10.42

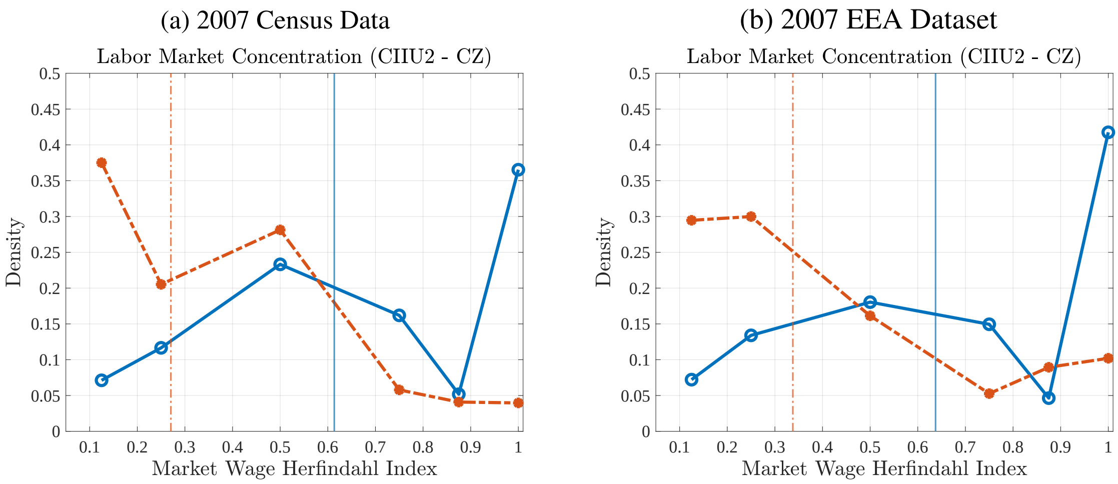
Notes. This table reports the percentage change in the average of selected outcomes across markets following the policy shocks discussed in Section 5.2.

Figure A.1: Correlation Between Employer Concentration Measures



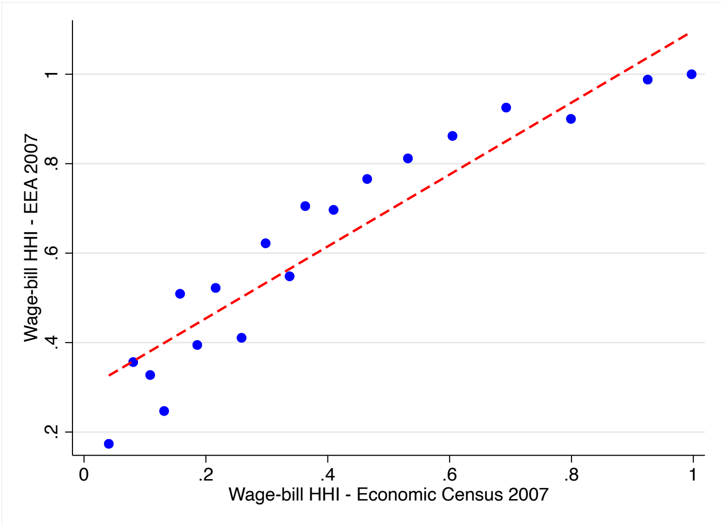
Notes. The figure plots the raw correlation of the three employer concentration measures – wage-bill HHI, employment HHI, and number of firms (bottom center panel) – one against the other across all local labor market-level observations. Wage-bill and employment HHI are strongly positively correlated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

Figure A.2: Employer Concentration Across Local Labor Markets in Census and EEA Data



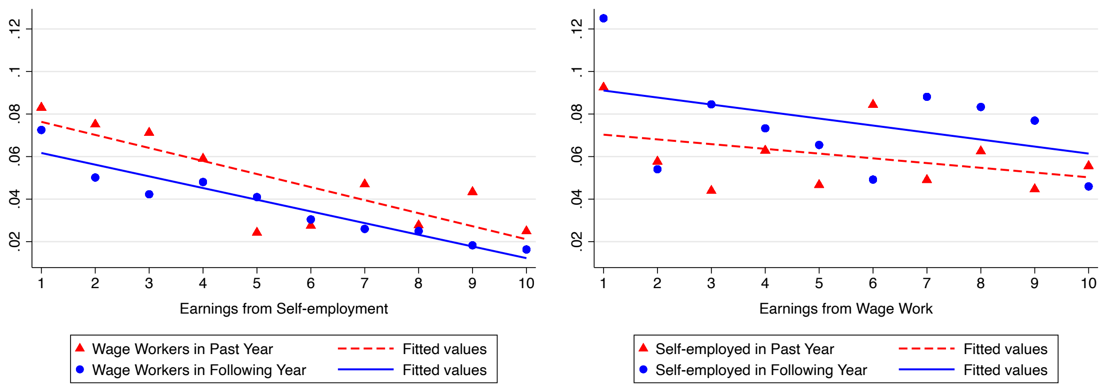
Notes. The figure plots the distribution of wage-bill HHI computed from the 2007 Peruvian Economic Census (left panel) and the same distribution computed from the 2007 EEA dataset (right panel) across local labor markets in the manufacturing sector. The blue solid line in both panels corresponds to the 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 A.3: Correlation Between Concentration Measures in Census and EEA Data



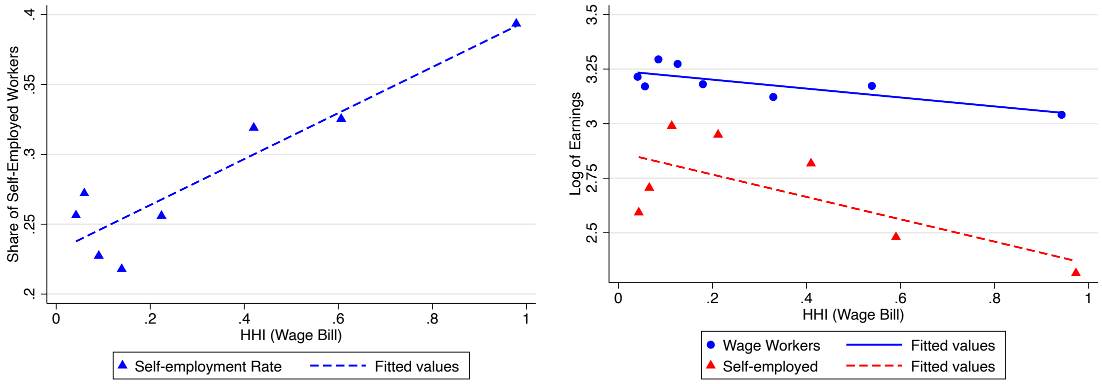
Notes. The figure illustrates the correlation between wage-bill HHI across local labor markets computed from the 2007 Peruvian Economic Census and the 2007 EEA dataset. Both variables are grouped into equal-sized bins, each point showing the average within bins. The dashed line shows the linear fit based on the underlying data, its slope equal to the corresponding parameter in the first column of Online Appendix Table A.2.

Figure A.4: Transition Probabilities and Earnings – Broadened Local Labor Markets



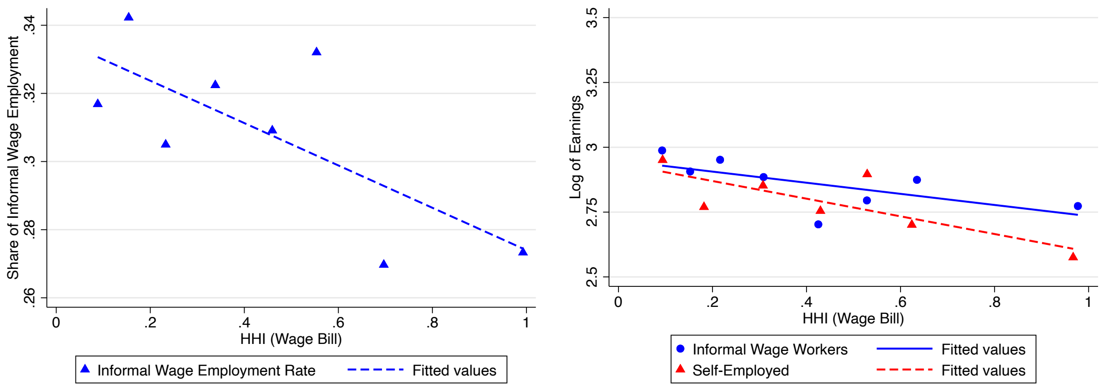
Notes. The figures illustrate the relationship between the likelihood of transitioning from and into wage work and self-employment, and earnings, where the latter are residuals from a regression of daily earnings over the full set of province or commuting zone fixed effects. The left panel plots average 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 A.5: Concentration, S.-E. Rate, and Earnings – Broadened Local Labor Markets



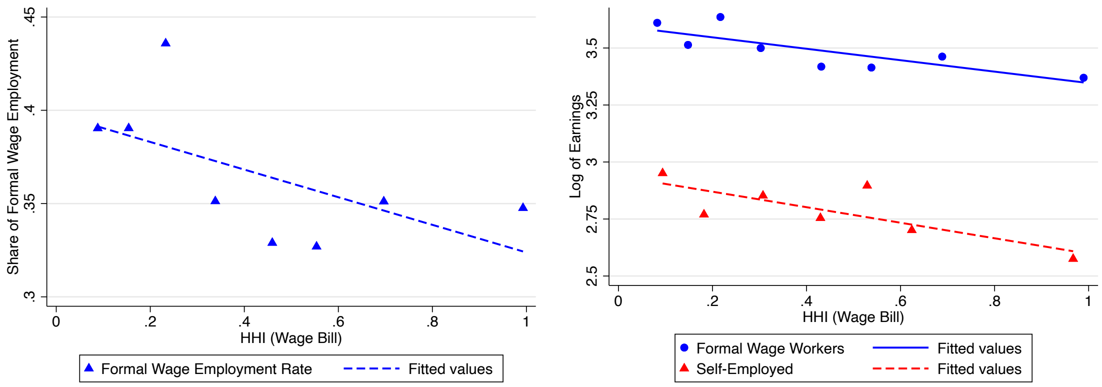
Notes. The figures illustrate the relationship between employer concentration, rate of self-employment (left), and earnings from both wage work and self-employment (right) across local labor markets as defined by 2-digit industries within departments. The left panel plots the share of self-employed workers 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.

Figure A.6: Concentration, Informal Wage Employment Rate, and Earnings



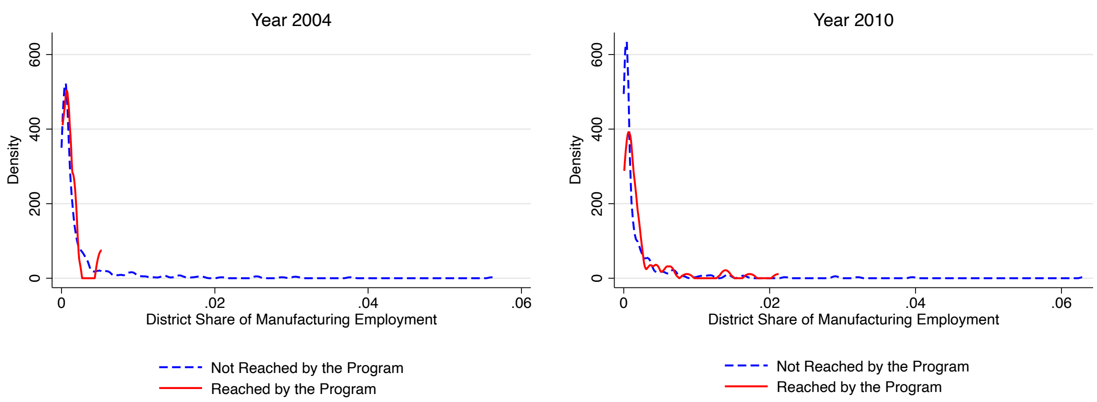
Notes. The figures illustrate the relationship between employer concentration, rate of informal wage employment (left), and earnings from both informal wage work and informal self-employment (right) across local labor markets. The left panel plots the share of informal wage workers in each decile of 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.7: Concentration, Formal Wage Employment Rate, and Earnings



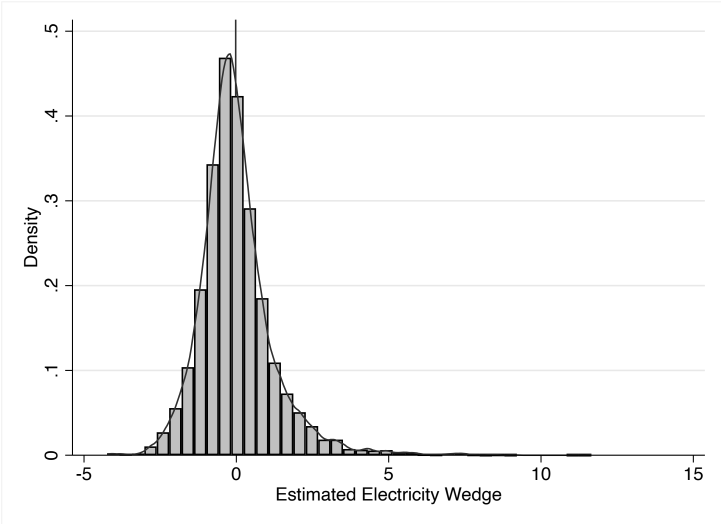
Notes. The figures illustrate the relationship between employer concentration, rate of formal wage employment (left), and earnings from both formal wage work and informal self-employment (right) across local labor markets. The left panel plots the share of formal wage workers in each decile of the wage-bill 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.8: Electrification Program Implementation Across Districts



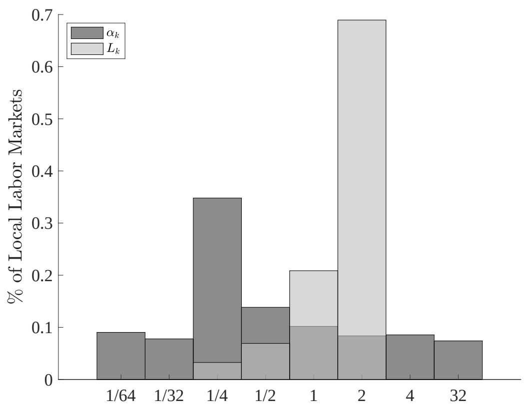
Notes. The figures shows the distribution of manufacturing employment share across districts reached vs. not reached by the electrification program in the first and last year of the IV estimation sample, i.e. 2004 and 2010.

Figure A.9: Distribution of the Estimated Electricity Wedge



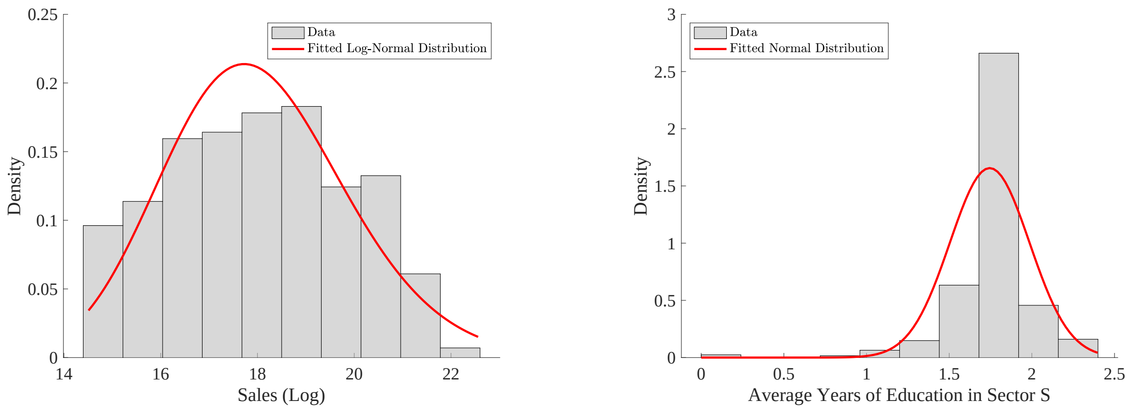
Notes. The figure plots the distribution of the estimated electricity wedges $\ln\widehat{(1+\tau^e_{ij})}$ obtained following equation (1) as the estimated residuals from the regression of the (log of) inverse electricity expenditure share $\ln\left(p_{ij}y_{ij}/e_{ij}\right)$ over the full set of 4-digit industry fixed effects and second-degree polynomials of output market shares in both the local labor market and economywide. The vertical bar indicates the value of the median across firms at baseline, i.e. as observed in the first year in which they appear in the data.

Figure A.10: 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.11: Sales and Education in the Data – Fitted Distributions

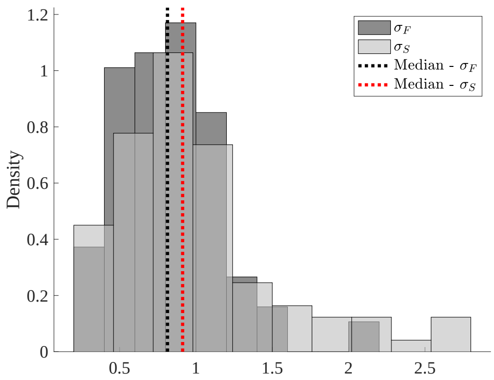


(a) Variance

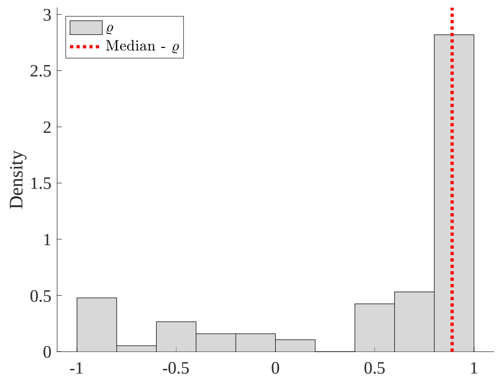
(b) Correlation

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.

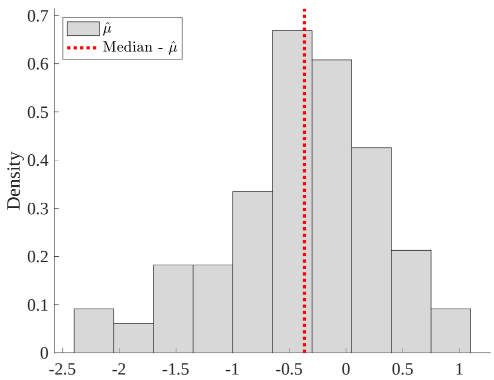
Figure A.12: Parameter Estimates for Joint Ability Distribution



(a) Variance



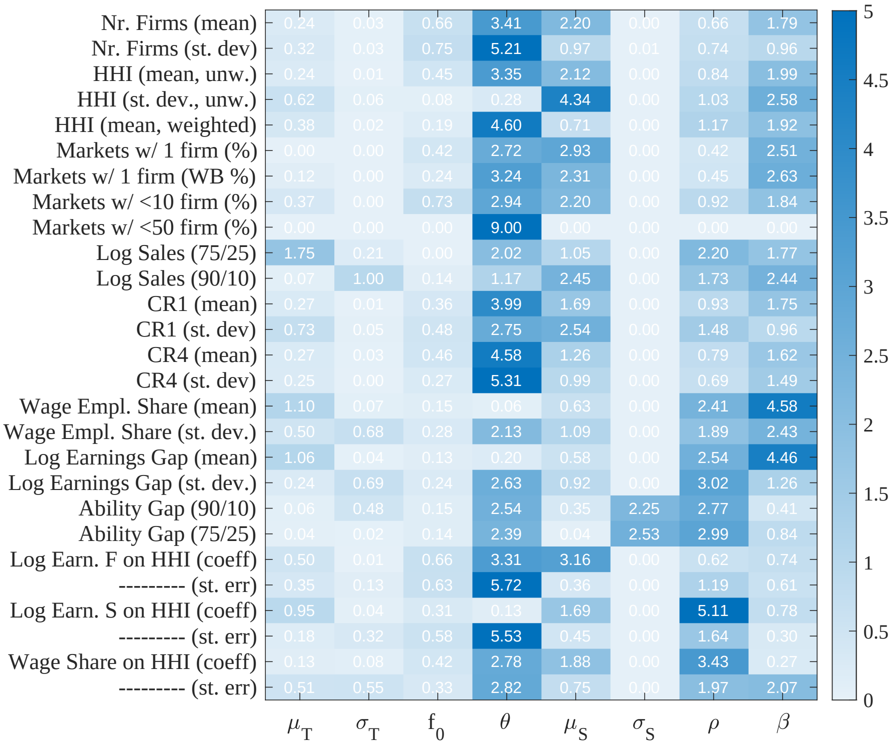
(b) Correlation



(c) Mean Comparative Advantage

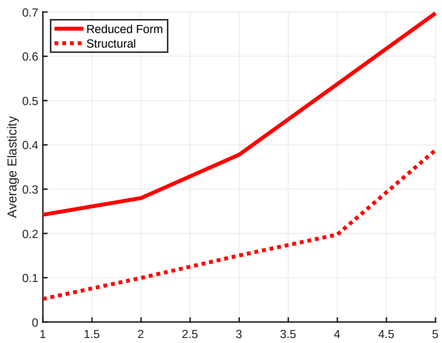
Notes. The figure shows histograms of the parameter estimates for the ability distribution, as discussed in Section 4.3. Panel (a) presents histograms for the estimates of $\sigma_{F,k}$ and $\sigma_{S,k}$ across local labor markets. Panel (b) displays the histograms for the parameter ϱ_k, while panel (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.13: Normalized Partial Derivatives of Moments with Respect to Parameters

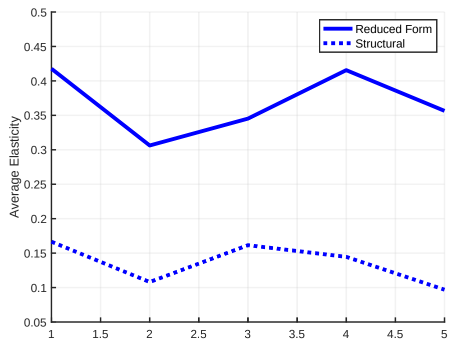


Notes. The Jacobian matrix includes the normalized values of the elasticity of each moment i with respect to a 10% change in parameter j around its estimated value while keeping all the other parameters constant. Each row is a moment and each column is a parameter.

Figure A.14: Inverse Elasticity, Concentration, and Self-Employment Shares



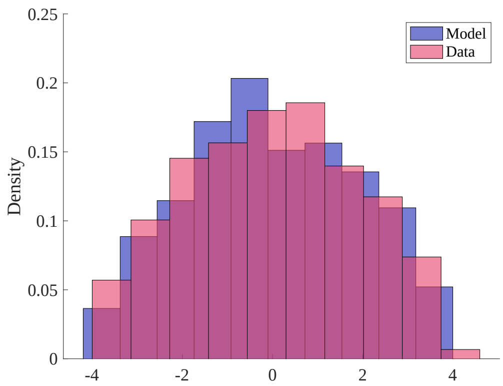
(a) HHI Quantiles



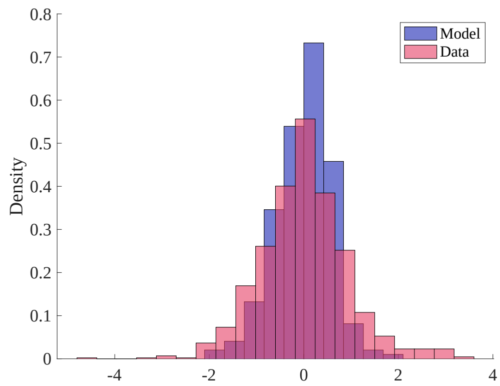
(b) Self-Employment Quantiles

Notes. The figure presents the average reduced-form and structural inverse elasticity of wage work 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.

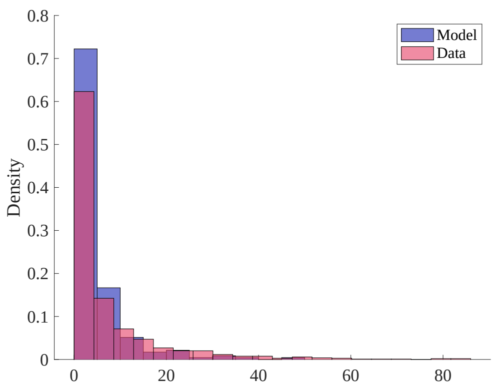
Figure A.15: Model Fit – Sales, Earnings Gap, and Number of Firms



(a) Sales (Log)



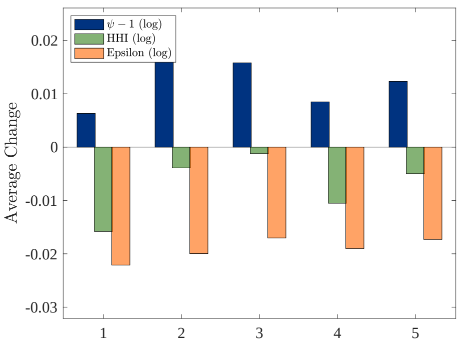
(b) Earnings Gap (Log)



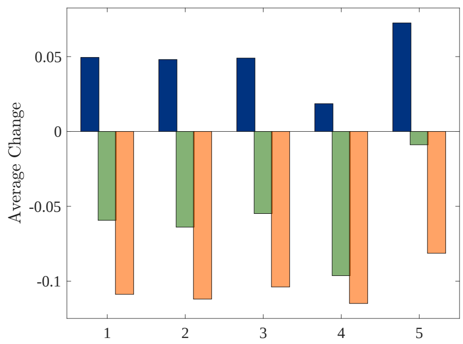
(c) Number of Firms

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

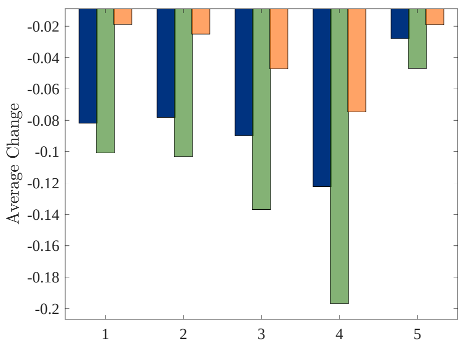
Figure A.16: Effect of Policy Shocks on Wage Markdown and Its Determinants



(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 work 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 Theory Appendix

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

B.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.

B.1.1 Market Equilibrium

Let $\boldsymbol{\Lambda}_k \equiv \{s_{iF,k}, s^N_{iF,k}, \mu_{iF,k}, \psi_{iF,k}\}_{i=1}^{M_k}$ 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 $\boldsymbol{\Lambda}_k = \left\{M_k, \hat{W}_k, \boldsymbol{\Lambda}_k\right\}_{k\in(0,1)}$ 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 $\left\{\{z_{iF,k}\}_{i=1}^{\bar{M}_k^*}, f_k^e, G_k\right\}$, all model equations (3.1)-(12) are satisfied.

Equilibrium given M_k Let's first assume that the number of entrants M_k is known in each k. From equations (7) and (8), we have:

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

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 (A.1) is equal to:

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

Using the pricing rule from (11) 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:

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

Equation (A.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 $\hat{W}_k$:

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

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

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

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 (A.4)-(A.6).
2. Given $\mathbf{\Lambda}_k$, the wage $\hat{W}_k$ solves equation (A.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:

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

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

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

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

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

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

B.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 (20). For simplicity, we focus on a single market k and omit the market-level subscript unless needed.

B.2.1 Aggregate Wage Work 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:

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

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

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

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:

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

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 (A.11) and (A.12) into equation (A.10) and taking the derivative with respect to $\ln \hat{W}$ we get the wage work supply elasticity:

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

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:

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

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 B.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 A.1. *When the joint ability distribution is log-normal and $\alpha \in (-1, 1)$, the aggregate supply elasticity of wage work $\epsilon_F(\hat{W})$ 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 (A.10)-(A.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 c_S, we have:

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

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

Proposition A.2. *When the joint ability distribution is log-normal $\alpha \in (-1, 1)$, the aggregate supply elasticity of wage work $\epsilon_F(\hat{W})$ is a one-to-one increasing function of the self-employment share. Conditional on the variance-covariance parameters, the self-employment share is a sufficient statistic for the given monotone decreasing function of the relative efficiency $\epsilon_F(\hat{W})$.*

B.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:

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

where $c_S = -c_F$ and all other parameters have been defined in Section B.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 (A.15), we can express the response of average ability in the two sectors as:

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

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 increase in the wage employment sector and decrease in self-employment as the relative wage $\hat{W}$ decreases.

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 decrease in both the wage employment and self-employment sectors as the relative wage decreases.

C Estimation Appendix

This section outlines the model's estimation strategy. Section C.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 C.2 covers the identification of remaining parameters using the Method of Simulated Moments. Section C.3 describes the computational algorithm used to solve the model.

C.1 Identification of Ability Distribution Parameters

C.1.1 Variance-Covariance Matrix

Here as well 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:

$$c_1 \equiv \Pr(h \in \text{wage sector}) = \Phi\left(\frac{\ln \hat{W} + \hat{\mu}}{\sigma^*}\right). \tag{A.17}$$

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:

$$\begin{aligned} c_3 &\equiv \mathbb{E}\left(\ln a_F W_F \mid a_F\hat{W} \geq a_S\right) = \ln W_F + \mu_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F), \\ c_4 &\equiv \mathbb{E}\left(\ln a_S W_S \mid a_F\hat{W} \leq a_S\right) = \ln W_S + \mu_S + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_S), \end{aligned} \tag{A.18}$$

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

$$\begin{aligned} c_5 &\equiv \text{Var}\left(\ln a_F W_F \mid a_F\hat{W} \geq a_S\right) = \sigma_F^2 + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2\left[-\lambda(c_F)c_F - \lambda^2(c_F)\right], \\ c_6 &\equiv \text{Var}\left(\ln a_S W_S \mid a_F\hat{W} \leq a_S\right) = \sigma_S^2 + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2\left[-\lambda(c_S)c_S - \lambda^2(c_S)\right], \end{aligned} \tag{A.19}$$

where c_F is as defined in equation (A.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 (A.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:

$$\begin{cases} c_1 & = \Phi(c_F), \\ c_3 - c_4 & = \sigma^* c_F + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F) - \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(-c_F), \\ c_5 & = \sigma_F^2 + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2 \left[\lambda(c_F)c_F - \lambda^2(c_F)\right], \\ c_6 & = \sigma_S^2 + \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)^2 \left[-\lambda(-c_F)c_F - \lambda^2(-c_F)\right]. \end{cases} \tag{A.20}$$

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 $\boldsymbol{\Theta}_k = (\sigma_{F,k}, \sigma_{S,k}, \varrho_k)$ in that market. We recover the vector $\boldsymbol{\Theta}_k = (\sigma_{F,k}, \sigma_{S,k}, \varrho_k)$ from a constrained Minimum Distance Estimation (MDE) procedure, where we restrict the variance coefficients to be non-negative and the correlation parameter to be $\varrho_k \in [-1, 1]$.

C.1.2 Mean Comparative Advantage

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

$$\begin{aligned} \mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) &\equiv \mathbb{E}\left(\ln a_F \mid a_F\hat{W} \geq a_S\right) - \mathbb{E}\left(\ln a_S \mid a_F\hat{W} \leq a_S\right) \\ &= \hat{\mu} + \left(\frac{\sigma_F^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_F) - \left(\frac{\sigma_S^2 - \varrho\sigma_F\sigma_S}{\sigma^*}\right)\lambda(c_S). \end{aligned} \tag{A.21}$$

While the terms involving the variance-covariance parameters are known, the left-hand side of (A.21) 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:

$$\mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) = \beta\mathbb{E}\left(\ln\frac{\text{edu}_F}{\text{edu}_S}\right) + \varepsilon. \tag{A.22}$$

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 A.22 can be decomposed as:

$$\mathbb{E}\left(\ln\frac{a_F}{a_S} \mid a_F\hat{W} \geq a_S\right) = \mathbb{E}\left(\ln\frac{N_F}{N_S} \mid a_F\hat{W} \geq a_S\right) - \mathbb{E}\left(\ln\frac{Em_F}{Em_S} \mid a_F\hat{W} \geq a_S\right), \tag{A.23}$$

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:

$$\mathbb{E}\left(\ln \frac{N_F}{N_S} \mid a_F \hat{W} \geq a_S\right) = \ln \frac{N_F}{N_S} - \frac{1}{2}\left(c_5 - c_6\right), \tag{A.24}$$

where c_5 and c_6 are the observed variances of log earnings in sectors F and S. Our model relates $\ln \frac{N_F}{N_S}$ to relative sector income adjusted for productivity:[1]

Combining these equations, we obtain:

$$\ln \frac{\text{Revenues}_F}{\text{Earnings}_S} = \alpha + \beta^* \mathbb{E}\left(\ln \frac{\bar{\text{edu}}_F}{\bar{\text{edu}}_S}\right) + \gamma_1 \left(\ln \frac{c_1}{1-c_1}\right) + \gamma_2 \left(c_5 - c_6\right) + u, \tag{A.25}$$

where $\beta^* = \beta \frac{\rho - 1}{\rho}$ and $u = \ln\left(Z_F\right) + e$, with e representing measurement error.

We estimate $\hat{\beta}$ through a regression of relative sector earnings on relative education, employment shares, and earning variances, adding controls for unobserved firm-level wedges in the right-hand-side.[2] 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 (A.22) and finally derive $\hat{\mu}$ using equation (A.21) market by market.

Robustness A potential concern with our approach is that the estimate of β could be biased if (i) there is measurement error in either the earnings data, or (ii) the orthogonality assumption is

[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, we can write:

$$\ln \frac{N_F}{N_S} = \text{const} + \frac{\rho}{\rho - 1} \ln\left(\frac{\text{Revenues}_F}{\text{Earnings}_S}\right) - \ln Z_F.$$

[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 the following regression specification:

$$\ln \frac{\text{Earnings}_F}{\text{Earnings}_S} = \alpha + \beta^* \mathbb{E}\left(\ln \frac{\bar{\text{edu}}_F}{\bar{\text{edu}}_S}\right) + \gamma_1 \left(\ln \frac{c_1}{1-c_1}\right) + \gamma_2 \left(c_5 - c_6\right) + v,$$

where the error term is now given by: $v = \ln\left(Z_F\right) + \ln \Phi_F + e$. This approach reduces measurement error by measuring the LHS variable 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 effect.

[3]For implementation, we use a conservative $\rho = 3$, which we later verify as close to the estimated value using the MSM method.

violated. To address this, we conduct a series of robustness tests. First, we verify that including controls for firms' Total Factor Revenue Productivity (TFRP) in regression (A.25) does not significantly affect the estimates of $\hat{\beta}^*$. We also explore an alternative specification for (A.25), where, instead of using $\ln \frac{\text{Earnings}_F}{\text{Earnings}_S}$ on the left-hand side, we use $\ln \frac{\text{Revenues}_F}{\text{Earnings}_S}$. Lastly, we consider the simple case of setting $\beta = 1$, such that $\mathbb{E}\left(\ln \frac{a_F}{a_S} \mid a_F \hat{W} \geq a_S\right) = \mathbb{E}\left(\ln \frac{\text{edu}_F}{\text{edu}_S}\right)$. While more restrictive, the latter approach is less susceptible to measurement error or omitted variable bias concerns.

C.1.3 Results

Figure A.12 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) illustrates the histogram for the correlation parameter $\hat{\varrho}$, and panel (c) depicts the distribution of the estimated mean comparative advantage $\hat{\mu}$ across markets. Although there is some heterogeneity among markets, the estimates of these parameters are generally well-behaved and centered near their mean values.

As explained in the main text, we set the values of the σ_k parameters and $\hat{\mu}$ constant across markets. Panel I of Table A.12 provides the median values for all the parameters. For $\hat{\mu}$, the table also includes estimates from the robustness checks discussed earlier. We choose these median values for calibration of the baseline model. This approach is justified for two main reasons. First, it greatly simplifies the computation of the model's equilibrium, thereby improving the efficiency of parameter estimation using the MSM method. Second, it aligns with our strategy of modeling heterogeneity, where a market in the data does not directly correspond to a market in the model. Instead, we estimate distributions of parameters from the data and treat the model's market as a multidimensional draw from this data-generating process. This issue is further compounded in the case of Roy estimates, as we constrain the variances to positive values and ϱ to lie between 0 and 1, leading to the exclusion of a few markets from Figure A.12.

C.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.13, 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.

C.3 Full Estimation Algorithm

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 $\mathbf{X} = (Y, 1)$.

[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.

(b) Given $\mathbf{X}$, we solve for the market equilibrium $\mathbf{K} = \{\mathbf{M}, \hat{\mathbf{W}}, \mathbf{\Lambda}\}$, as detailed in Appendix B.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 $\hat{\mu}$.

(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 50,000 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 10,000 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, $\hat{\Phi} = \arg\min_{\Phi} \mathcal{L}(\Phi)$.

C.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

[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.

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. 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:

$$\ln z_{i(j,g)t} = \gamma PER_{gt} \times EC_{i(j,g)} + \phi_i + \delta_{(j,g)t} + v_{i(j,g)t}, \tag{A.26}$$

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.