Title

Applying Causal Inference to Topics in Labor and Development Economics

ISBN

9798288861901

Author

Wang, Zicheng

Publication Date

2025-05-15

Applying Causal Inference to Topics in Labor and Development Economics

by

Samuel Zicheng Wang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Benjamin Faber, Chair
Professor Edward Miguel
Professor Andrés Rodríguez-Clare

Spring 2025

Applying Causal Inference to Topics in Labor and Development Economics

Abstract

Applying Causal Inference to Topics in Labor and Development Economics

by

Samuel Zicheng Wang

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Benjamin Faber, Chair

The central theme of this dissertation is applying causal inference methods to the analysis of issues of policy importance in labor and development economics, using knowledge in econometrics and quantitative economics learned throughout the PhD. Recent advances in causal inference—rigorously studying whether and how changes in certain factors affect the outcome—have greatly enriched econometrics and economics. In this work I apply several methods to study policy questions that affect businesses, workers, and consumers worldwide across a variety of contexts to demonstrate their applications to real-world economics issues.

Chapter 1 of this dissertation applies a shift-share instrumental variables (SSIV) approach to study the effect of labor demand shocks on local consumer prices in Mexico. Policymakers have been interested in how local labor demand shocks affect welfare. While many studies have examined the effects of local shocks on nominal income, much less attention has been paid to prices, the denominator of real income. Using data on the universe of CPI price quotes, workers, and firms in Mexico, I study the effects of local labor demand shocks on local consumer prices. To base the estimation on plausibly exogenous variation, I propose an Instrumental Variables strategy that leverages national industry-level changes in labor demand in combination with pre-existing employment shares. I find that a 1 percentage point increase in the growth rate of local labor demand leads to a 0.061 percentage point increase in services inflation, but a 0.099 percentage point decrease in goods inflation. The effects are concentrated in products that exhibit larger geographic price variation and are more subject to local markups. The decrease in product prices is consistent with increased product entry as firms introduce new varieties in locations with increasing market size, while services are more exposed to increases in local wages. To guide the interpretation of empirical results, I develop a general equilibrium model that features variable markups and endogenous firm entry. Comparative statics show that an increase in market size driven by local labor demand shocks leads to relatively lower markups and prices on continuing varieties. The overall effect on non-housing inflation is close to zero as lower inflation in tradables cancels

out higher inflation in services, but poorer households see their price indices fall more than richer households, who spend more on services.

Chapter 2 of this dissertation, joint with Uyanga Bambaa, Edward Miguel, and Michael Walker, estimates the gender wage gap and examines the underlying factors that cause the gap in Kenya. The gender wage gap remains persistent across the world, especially in low- and middle-income countries. We estimate the magnitude of the gender wage gap and contributors to it in Kenya utilizing a rich dataset from the Kenya Life Panel Survey (KLPS) covering 5,878 adults. The data measure typically unobservable individual characteristics that could potentially determine wages, such as cognition, personality traits, job task indices, and economic preferences, allowing for estimation of the gender wage gap among prime-age adult workers in Kenya controlling for extensive covariates. We use both regression and causal machine learning methods to estimate the gender wage gap. We find that women earn 78 log points less than men (54%) without adjustments, and 41 log points less (34%) after controlling for education, experience, demographics, occupation, cognition, personality, preferences, and job task indices. The latter four account for 18% of the residual gap unexplained by education and occupation. The large gender wage gap after accounting for a rich set of potential confounders potentially suggests the role of discrimination and social norms in hindering women's access to opportunities in the labor market.

Chapter 3 of this dissertation, joint with Matthew Grant and Meredith Startz, combines causal inference with structural modeling in a setting where rigorous identification is challenging to yield insights on the gains from firm consolidation among small retailers in Nigeria. We ask: Why are consumer goods in low-income countries often sold in a physical market area with many small firms side-by-side, rather than by a large, integrated retailer as is more common in high-income countries? To what extent does this matter for prices? We begin by documenting a set of novel empirical patterns about markets in Lagos, Nigeria, using original survey data among 1,500 firms. We show that wholesale/retail firms incur large fixed costs when sourcing goods for resale from distant suppliers; as a result, they often source from very nearby suppliers within their own market. We also document that labor and supervision costs are increasing in the number of employees, consistent with span of control or agency problems. We build a structural model in which these countervailing economies and diseconomies of scale lead to a large number of co-located, independent wholesalers and retailers who transact with one another but maintain boundaries between firms as the equilibrium market structure. Compared to standard models accounting for the firm size distribution in developing countries, this leads to substantively different conclusions about the welfare costs of small firms. Quantifying this comparison is still in progress, but the qualitative insight is that small firms that co-locate and source within a market capture some of the gains from scale in sourcing while aligning labor incentives. In essence, there is less gain to be had from turning micro-entrepreneurs in the retail sector in a Nigerian market into cashiers in a large firm than might be expected.

To My Family and Friends

Contents

List of Figures

List of Tables

Acknowledgments

I am deeply grateful to my dissertation chair Benjamin Faber, who mentored me throughout my doctoral studies from the very beginning and encouraged me to pursue big, meaningful questions. Completing this dissertation would not have been possible without his guidance. I am also incredibly grateful to my committee members, Edward Miguel and Andrés Rodríguez-Clare, for their insightful advice, encouragement, and support. Beyond my advisors, I owe a major debt to additional faculty members at Berkeley. I would like to in particular thank Matilde Bombardini, Marco González-Navarro, Cécile Gaubert, Supreet Kaur, Jeremy Magruder, Nick Tsivanidis, and Christopher Walters for their guidance. I would also like to thank my pre-doctoral mentor Amy Finkelstein, who introduced me to the wonderful world of economics research.

I also want to thank my collaborators. I am grateful to Edward Miguel, Michael Walker and Uyanga Bambaa for the collaboration on the Kenya labor project. I am indebted to Meredith Startz for introducing me to the Lagos project and supporting my research throughout my PhD journey. Matthew Grant, thank you for being patient in guiding me through the theory work. I thank the National Institute of Statistics and Geography (INEGI) for access to the prices and census data analyzed in Chapter 1.

I am fortunate to have developed lasting relationships with other students in the PhD program. In particular, I would like to thank Matteo Saccarola, Grady Killeen, Richard Jin, Alfredo Mendoza Fernández, Lucy Hackett, Bailey Palmer, Nadav Tadelis, Samuel Ayers, and Catherine Che. I also want to thank David Wu, Luisa Cefalà, and Nicholas Swanson, whose mentorship has been instrumental in my PhD journey.

I also wish to thank my non-Berkeley Economics friends in the Bay Area and beyond, without whom I could not have finished this journey. I am deeply grateful for your presence in my life.

Chapter 1

Effect of Labor Demand Shocks on Consumer Prices

1.1 Introduction

How do local shocks affect welfare? Economists have long examined the welfare implications of shocks to local labor markets. A large literature has examined the effects of adverse local labor market shocks, such as recessions (Yagan, 2019) and import competition from China (Autor, Dorn, and Hanson, 2013), or positive shocks such as the arrival of manufacturing plants (Greenstone, Hornbeck, and Moretti, 2010) and regional development policies (Kline and Moretti, 2014) on workers' employment and nominal wages. However, much less is known about the effects on *real* income—which is the true object of welfare calculations—because consumers value the purchasing power of their income and not just their nominal income. An increase in local labor demand increases employment and nominal wages, which in turn may push up the price index by increasing product demand and thus attenuate welfare gains for consumers. Examining the effects on real income necessarily requires detailed data on changes in local prices, which has been a daunting task due to the lack of available data on price changes. Some studies, primarily in a US context, have examined housing prices, yet little is known about the effects of local shocks on tradable and non-housing nontradable prices.

Leveraging rich barcode-by-store-level Consumer Price Index (CPI) microdata across Mexican municipalities between 1993 and 2018, I quantify the effects of quasi-exogenous changes to local nominal income driven by local labor demand changes on local price changes. I use a Bartik (1991)-style shift-share instrument that combines industry-level baseline employment shares in each municipality with national changes in industry employment in order to isolate changes to local labor demand that are uncorrelated with location-specific shocks to prices. I combine a rich set of publicly available barcode-by-store-level price quotes for more than 85,000 items across 315 product categories, which constitute the source data of the Consumer Price Index in Mexico, with confidential establishment-level data from the

Economic Census to construct five-year changes in prices of the same barcode-by-store-level item across municipalities in Mexico. By examining price changes for the same item sold in the same place, I prevent confounding from unobserved changes in product quality or product attributes, which has been a common challenge in the literature that measures inflation. I then regress those price changes on both the instrument and on changes in local income in the reduced-form and Instrumental Variables (IV) specifications respectively. The identifying assumption of the shift-share instrument relies on the exogeneity of national industry-level shocks following Borusyak, Hull, and Jaravel (2022), which I validate with several balance tests and by checking for pre-trends.

I show that, in the reduced-form specification, a 1 percentage point increase in the growth rate of local labor demand leads to a 0.061 percentage point increase in the services price inflation rate, but a 0.099 percentage point *decrease* in the goods price inflation rate. In the IV specification, a 1 percent increase in the growth rate of local income leads to a 0.12 percentage point increase in services price inflation, but a 0.18 percentage point in goods price inflation. I find that the effects are driven by decreases in the prices of unprocessed food, clothing, transport, recreation, and personal care goods, which exhibit higher price dispersion across space and tend to be priced more locally instead of following national pricing rules. To explore the mechanisms that drive these perhaps counterintuitive results, I perform several tests that suggest that as local income grows, firms introduce a larger number of new varieties. This leads to a higher number of substitutes and exerts competitive pressure on existing varieties.

To guide interpretation of the results with economic theory, I develop a tractable general equilibrium model that features endogenous markups and falling demand elasticity as the number of varieties increases. Comparative statics show that in response to an increase in market size, firms introduce more varieties, which drives down the markup on existing varieties. This is similar to the mechanism in Jaravel (2019), who finds that the prices of existing varieties rise less quickly for households that experienced faster income growth in the US as a result of firms' targeting product innovation toward high income-growth consumer segments. In contrast, services are more exposed to increases in local wages, which leads to different dynamics. Finally, I examine the overall non-housing price index implication by combining location and income quantile-specific expenditure shares with my estimates. I exclude housing from my empirical analysis because housing costs are measured with substantial noise in the CPI microdata. I find that an average household experiences little change in overall inflation following an increase in the growth rate of local labor demand, but positive local labor demand shocks tend to lower prices for the lowest-income households and raise prices for the highest-income households, because poorer households spend a higher share of income on goods and less on services. This points to the potential for positive local labor market shocks to reduce within-location inequality between richer and poorer households by slowing inflation for products that poorer households consume more of.

The analysis proceeds in four steps. In step 1, I document a set of stylized facts about price dispersion across Mexican municipalities. I show that across Mexican cities, there is large variation in the consumer price index: The 90th percentile metropolitan area is 33%

more expensive than the 10th percentile metropolitan area. Such geographic heterogeneity persists even after excluding housing prices, which contrasts with patterns in US data from Diamond and Moretti (2021), who document that the vast majority of spatial variation in the price index can be explained by variation in housing costs. I also show that, at product category level, tradable product prices are 10%-25% percent higher at the 75th percentile than at the median. These results demonstrate significant variation in prices across locations within the same product category: The spatial distribution of prices in the Mexican economy looks very different from that in the US—where, for example, DellaVigna and Gentzkow (2019) show that national big-box retailers set uniform prices for the same item across locations. In my context, price dispersion and differences in prices across locations suggest the presence of variable markups for tradable goods across locations within a middle-income country experiencing rapid and unequal growth, similar to Simonovska (2015), who documents variable markups for identical tradable goods across countries with varying income levels.

In step 2, I perform my main empirical analysis. To test the validity of the shift-share IV, I follow Borusyak, Hull, and Jaravel (2022) and show that the *shocks* in the shift-share can be argued to be as good as randomly assigned. Since my causal model is in first differences, identification rests on a 'parallel trends-style argument in which industry-level shocks—employment changes in my case—are orthogonal to changes in industry-level unobservables, conditional on industries' exposure shares. I perform balance tests on pre-trends, i.e., pre-period price changes, and pre-exposure industry-level variables to show that shocks can be assumed to be as good as randomly assigned. I then perform reduced-form and IV regressions using five-year differences in the price of the same item (i.e., a barcode-by-store level product), and show that an increase in the growth rate of local labor demand leads to higher services inflation but lower tradable goods inflation. The results are robust to controlling for a rich set of pre-exposure location characteristics variables, potential changes in trade costs, region-specific time trends, and direct productivity effects that lower the cost of production.

In step 3, I explore the mechanisms that drive the main findings. The main results are perhaps surprising to economic intuition: As a positive local labor demand shock increases local income, local demand for goods increases, which should drive up prices in equilibrium. How could a positive local income shock decrease tradable prices? Price equals markup times marginal cost, so the observed decreases could come from a reduction in marginal cost, a reduction in markup, or both. A reduction in marginal cost may be possible if there are external economies of scale at retail level. Some earlier studies have provided suggestive evidence of increasing returns to scale at the sector level in retail (Ofer, 1973; Kato, 2012). However, I do not find that predicted local labor demand increases leads to any increase in local retail employment, which would be a necessary condition for the external economies of scale channel. On the other hand, positive income shocks could lead to pro-competitive pressure from newly introduced varieties if firms endogenously respond to the growth in market size by targeting product entry to booming locations. A decrease in markup may be possible if richer consumers switch out of existing varieties in favor of newly introduced, more expensive varieties.

Although I do not observe the entry of new varieties directly, I provide several pieces of evidence in support of this mechanism. First, the main results only include products that have never been substituted across the five-year period to rule out potential changes in product characteristics that are correlated with price changes. Interestingly, when I perform the same analysis on the sample of items that experienced a change in product characteristics but are still the same kind of product (e.g., changing from a pine wood bed to a poplar wood bed, which was a newly introduced model), results for both tradables and nontradables are positive. This suggests that the price evolution of products that experience innovation follows a different trajectory from those that have consistently had the same product attributes, and thus does not favor households in growing locations. Second, I examine the effect of predicted local demand changes on the number of item substitutions in the CPI catalog, which is an indication of the degree of new product entry to local markets because the statistical authorities (Banco de México/INEGI) determines item substitutions to maintain the representativeness of the CPI consumption basket. I find that a 1 percentage point increase in predicted local labor demand growth leads to a 0.25 percentage point increase in the share of items that experience product substitution, which suggests that booming locations see a higher share of new varieties being introduced. Third, I regress changes in the number of local retail establishments on predicted changes in local labor demand. A 1 percentage point increase in predicted local labor demand growth is associated with a 1 percentage point increase in the growth rate of department stores within a municipality. This shows that places experiencing high growth have also attracted modern big-box retailers, which tend to offer a wider range of available varieties, and thus likely reinforce the pro-competitive effects on the prices of existing items.[1] Effects on all other types of retail establishments are not statistically different from zero.

For services, the results are different. If variation from the shift-share IV is valid, then we would expect services prices to increase since the pro-competitive effects due to product entry are specific to products, whereas the primary cost of services is local labor, which becomes more expensive following a labor demand shock. This is exactly what I see: The shock has a positive, albeit borderline statistically significant, effect on services inflation. Beyond the fact that services are more exposed to increases in local wages, services products are also likely to be less substitutable for each other compared with tradable goods. Establishments enjoy market power due to product differentiation and personalized relationships between consumers and firms, which leaves less room for pro-competitive effects.

In step 4, to guide interpretation of the empirical results, I write down a tractable general equilibrium model that features non-homothetic preferences and variable markups consistent with an increase in local income that leads to higher product variety and lower relative prices for existing varieties. The model builds on Feenstra (2003) and Jaravel (2019), but adds location-specific markups and costs of production. The model features non-homothetic CES

[1]Modern retail establishments such as department stores and big-box retailers may also have lower marginal costs and thus lower prices directly. However, since my main sample includes barcode-by-store items that were present continuously during a five-year period, the entry of modern retailers does not contribute to my main estimated effect.

preferences across sectors and translog preferences across varieties within a sector. Firms are homogeneous within sectors and make specific decisions to enter each location or not according to its market size. Market size is larger if there are more consumers or if each consumer's purchasing power is higher. I derive comparative statics that show that the equilibrium number of varieties in a sector is increasing in market size. In my setting, an increase in local labor demand increases market size by increasing local income. As a result, firms introduce a larger number of new varieties to those booming locations. As the number of varieties increases, the number of substitutes for each varieties increases from the perspective of the consumer, and the elasticity of demand for each variety increases. In turn, this decreases the markup on any given variety and thus leads to the lower tradables inflation in growing markets observed in the empirical analysis.

Finally, I examine the effects of local labor demand shocks on real income by combining location and income quantile-specific expenditure shares with product group-specific estimates of the effects of a local labor demand shock into a a simple Laspeyres index. I conduct a back-of=the-envelope calculation to estimate the price effects of a 10 percentage point increase in the growth rate of local labor demand—which corresponds to moving from the median to the 90th percentile in the sample—for households across cities and the income distribution. I find that the average effect on non-housing inflation is close to zero, but this masks heterogeneity across the income distribution. Given a 10 percentage point increase in the growth rate of local labor demand, the richest 20% of households in each metropolitan area would experience a 0.084 percentage point increase in the non-housing inflation rate, but the poorest 20% of households would experience a 0.154 percentage point decrease in the non-housing inflation rate. This is because low-income households tend to spend a lower share of their income on services and a higher share on goods, thereby benefiting more from the pro-competitive effect of falling goods inflation as a result of the positive local labor demand shock.

This paper relates to several strands of literature. First, my findings contribute to the longstanding literature in labor economics (Yagan, 2019; Autor, Dorn, and Hanson, 2013; Kline and Moretti, 2014; Greenstone, Hornbeck, and Moretti, 2010) and development economics (Egger et al., 2022; Bastagli et al., 2016) that studies the effect of local income shocks on local economic outcomes and welfare. This includes both direct shocks (through cash transfers) and indirect shocks (through a positive increase in local labor demand), such as place-based policies or differential exposure to aggregate economic trends. Relative to the literature that typically only examines effects on nominal income, this study exploits high-quality comprehensive administrative data on prices to study the effect of local shocks on consumer welfare in real terms. Exceptions in the US literature focus on housing, such as Diamond (2016), Hornbeck and Moretti (2022), and Notowidigdo (2020). This paper aims to expand the literature by studying the effects of demand shocks on non-housing prices, which together constitute a larger share of the consumer budget than housing, particularly in a developing country such as Mexico. In the development literature, Cunha, De Giorgi, and Jayachandran (2019) study the effect of a cash transfer program in ultra-poor villages in Mexico, and find that in-kind transfers lower prices in treatment villages. Egger et al.

(2022) study a large-scale cash transfer program in western Kenya and find almost zero effect on local prices, which they attribute to an increase in factor utilization and a decrease in the "slack capacity" of labor. Relative to those papers, this study examines municipalities across a large country of primarily urban consumers across the income distribution, and demonstrates that in more complete and developed markets, local income shocks can even lead to negative effects on local goods prices.

This paper also relates to an emerging literature on real income inequality across space and the income distribution. A recent set of papers (Moretti, 2013; Diamond, 2016) have documented differences in the cost of living across US cities that contribute to real wage inequality; a related literature in macroeconomics has also documented inflation inequality among households across the income distribution (Jaravel, 2019; Jaravel, 2021; Faber and Fally, 2022). This paper highlights how shocks to local labor demand and income could exacerbate spatial real income inequality through their effects on prices, in addition to effects on nominal income. In my context, households in cities that experience a positive income shock also see their goods prices decline, which can exacerbate spatial real income inequality. However, within a city, richer households consume disproportionately more services products, which causes their price index to rise faster than that of poorer households because the shock increases services inflation while decreasing goods inflation. Thus, a positive local labor market shock has the potential to reduce inequality within locations.

Furthermore, this paper speaks to the literature in international trade that studies the effects of openness to trade and globalization on price indices and variable markups (Faber, 2014; Simonovska, 2015; Fajgelbaum and Khandelwal, 2016). Although I do not directly examine a trade shock, this paper studies how shocks specific to local economies translate to effects on prices. During my study period, Mexico experienced rapid trade liberalization accompanied by structural transformation: Mexico joined the North American Free Trade Agreement (NAFTA) on January 1994—three months after the beginning of my study period—and employment in services jumped from 52% in 1993 to 62% in 2018. In my setting, trade liberalization and structural change constitute some of the forces behind the increase in local labor demand across Mexican municipalities. These forces, through increasing local income and providing incentives for manufacturers and retailers to introduce a larger number of varieties, likely lead to a fall in the markups of existing varieties. Richer municipalities not only see a higher rate of entry of new varieties, but also slower price growth of tradables due to the increase in income. This highlights the importance of studying the price effects of shocks to nominal income through the markup channel.[2]

[2] One caveat of my results is that since I am comparing municipalities with large vs small labor demand shocks, I am not capturing the full general equilibrium effects. Municipalities experiencing a small local labor demand shock may still experience a decline in tradables prices compared to the counterfactual with no labor demand changes, just less so than those with a large shock.

1.2 Data

CPI Microdata

Monthly price quotes data are from store price surveys from 1993 to 2018, which were conducted to calculate the Consumer Price Index (CPI) by Banco de México, Mexico's Central Bank, until June 2011 and the National Institute of Statistics and Geography (*Instituto Nacional de Estadística, Geografía e Informática, INEGI*) from July 2011 onward. These monthly price quotes are published every month in the Diario Oficial de la Federación, the official government gazette. The sampling of price quotes is designed to capture a representative sample of Mexican household consumption in tradables (unprocessed food, processed/packaged food, non-food goods items), services, and housing across various establishments (supermarkets, open markets, convenience stores, street vendors, etc.) in 46 metropolitan areas and 128 municipalities. The number of price quotes collected in each municipality is determined based on market size, in order to yield a representative sample that captures the price index faced by a representative Mexican consumer. Each month of data contains approximately 30,000 individual price quotes until 1995, after which the number of price quotes expanded to more than 50,000.[3] Any changes in the presentation, appearance, size, modality, or sourcing location of a price quote are reported as an item substitution in the official government gazette (Gagnon, 2009; Faber, 2014), which enables me to identify identical items over time without changes in sourcing location or product characteristics. A subset of goods items is identifiable at barcode level (e.g., Nestle condensed milk, La Lechera, can of 397 grams; Matches, Talisman brand, box of 55 units; Bimex brand bicycle, Intrepid 2000, Model 705), but even for raw produce items that do not contain a brand description, the surveys record the price of the same item (e.g., Beef, American cut, New York strip, 1Kg) at the same store every month, so the price series I construct is as close to barcode-identifiable as possible and should address any confounds from changes in product characteristics. For services items such as doctor visits, haircuts, or tourist packages, it is not possible to condition on the exact quality and attributes, though CPI enumerators collect the prices of identical services (e.g., a 4-day hospital stay with room and board at the same hospital or a chicken parmesan sold over lunch at the same lunch counter) over time to make the prices comparable. Appendix A documents additional details about the CPI microdata.

To compute price changes that correspond to the years of the Economic Census, which takes place every five years, I construct 3-month averages of Q4 1993, Q1 1998, Q4 2003, Q1 2008, Q4 2013, and Q1 2018. Selection of a 3-month window at the end or beginning of the year is designed to include as many items as possible that had not experienced substitution—i.e., a change in product characteristics, in which case Banco de México/INEGI determines that another variety, perhaps a newer one, is more representative of the consumption basket and replaces the old variety with the new one. In my main analysis I restrict attention to

[3]For product groups in food and beverages the reported monthly price quotes are averages across 2-4 monthly price quotes for each item. All other price quotes are collected monthly (Faber, 2014).

products that had never been substituted across the five-year window in order to exclude any confounds from product characteristic changes. In a separate exercise, however, I also include in the sample items that had experienced substitution. Two updates to the sampling methodology and baskets occurred in 2002 and 2010, during which there was an overhaul of the item codes used and thus I was not able to match item codes between pre- and post-update years. This means that I was able to construct three disjoint five-year differences for my analysis: 1993-1998, 2003-2008, and 2013-2018.

Using the confidential list of store names and municipality identifiers for all price quotes, I identify the municipality of each price quote starting from 2002. For price quotes prior to 2002, I can only identify the metropolitan area of the price quote in the data, because INEGI did not have the municipality identifiers for those years. Thus, for the 1993-1998 difference I construct local employment and wage changes at metropolitan area level; for the 2003-2008 and 2013-2018 differences I construct local employment and wage changes at municipality level. There are 46 metropolitan areas and 128 municipalities covered by the CPI microdata. Figure 1.1 shows their geographic distribution across Mexico. The data cover all major Mexican metropolitan areas and include municipalities in each of the regions of Mexico.

The CPI data also include official weights for each of the 315 product categories used to calculate the CPI. These weights are constructed from household expenditure surveys, updated every 8-10 years, and used by authorities to weight price changes across product categories. I use those weights in my regressions to ensure that I estimate the effects on a representative price quote in the typical consumer's consumption basket. Housing prices are noisily measured; as a result, in my empirical analysis I do not include housing price quotes.

Economic and Population Census

The Economic Census is a complete enumeration of all fixed establishments in Mexico except for street vendors every five years (Hsieh and Klenow, 2014). I use confidential microdata for the Economic Census from 1993 to 2018 to construct municipality-level measures of employment and wages, as well as predicted local labor demand changes through the shift-share instrument, by combining national changes in industry employment with municipality-level industry employment shares. I divide the total wage bill per municipality by the number of workers to obtain the per capita measure of wages in each municipality. In a robustness exercise, I also use the number of adults as an alternative measure, interpolated linearly from the decennial Population Census figures for each Economic Census year as the deflator, to account for potential compositional changes in the labor force: For example, if low-skilled workers primarily exit during the recession, wages per worker will be artificially inflated, but that does not reflect the true health of a local labor market.

Household expenditure survey data

I use the National Household Income and Expenditure Surveys (*Encuesta Nacional de Ingresos y Gastos de los Hogares*, ENIGH), which are administered biannually by INEGI between 1994 and 2012 to a nationally representative sample, to construct the housing prices and characteristics used in the descriptive exercise in the next section. ENIGH contains data on household rent (or owner's equivalent rent, if the household does not pay rent)[4] over a three-month period, as well as a set of housing characteristics. These include the year the housing structure was built, number of rooms, number of bedrooms, presence of bathroom(s), access to water, etc. I use these variables to predict city-level rent conditional on housing characteristics in the exercise to construct spatial price indices in Section 1.3.

I also use the ENIGH survey to compute households' expenditure shares across product categories. ENIGH captures households' expenditure shares across all retail and non-retail product groups that match those in the CPI data. Appendix A provides additional details on the data.

1.3 Descriptive evidence

Price dispersion across cities

How dispersed are prices across Mexican cities? In the US, retailers tend to set uniform prices for packaged goods across different locations, which decreases spatial variation in prices. If prices of tradables are set nationally, there may be little room for location-specific shocks and policies to affect the prices of goods sold in one location over others. However, in a middle-income country such as Mexico, large disparities in income and growth across locations persist, and barriers to domestic trade are high. Markets are also less integrated than those in developed countries: Instead of being dominated by a few large retailers, domestic and local small-scale retailers play a large role in the retail market because Mexico only began seeing substantial expansion of large multinational retail chains in the early 2000s (Atkin, Faber, and Gonzalez-Navarro, 2018). Pricing by small, local retailers likely leads to differences in markups for the same or similar items across locations, since each retailer sets prices according to location-specific consumer demand and thus contributes to spatial price dispersion.

In this section I document large differences in prices across locations in Mexico. As in Diamond and Moretti (2021), to construct spatial price indices, I closely follow the methodology the Bureau of Labor Statistics (BLS) uses to construct CPI in the US, but generalize it to capture variation across metropolitan areas in Mexico. The BLS uses a Laspeyres price index to calculate the CPI in the US; in my setting, I construct the following spatial

[4] Precisely, "estimated value of rent" (*valor estimado del alquiler*).

Laspeyres price index for metropolitan area j:

$$P_j^{Laspeyres} = \sum_{i \in I} \frac{p_{ij}}{\overline{p_i}} s_i \tag{1.1}$$

where i indexes one of the 315 product categories. p_{ij} is the average price of items within product category i and metropolitan area j. $\overline{p_i}$ is the average price of items within product category i in the reference metropolitan area, and s_i is the nationwide share of consumer expenditure on product category i, obtained from Banco de México/INEGI from their CPI calculation. Consistent with the official method employed by Banco de México/INEGI, the index is aggregated from product category level.[5] I choose Guadalajara to be the reference metropolitan area because its monthly rent for a given vector of housing characteristics is roughly equal to the median rent in the analysis sample. As a result, the normalization implies that the price index of Guadalajara is equal to 1 and the indices of other cities are to be interpreted relative to Guadalajara at a given point in time.

To measure housing costs, I again follow Diamond and Moretti (2021) to compute predicted monthly rent at city level given a set of housing characteristics, which accounts for differences in the type and quality of housing across locations. I estimate a household-level hedonic model using ENIGH expenditure survey data and regress three-month rent on metropolitan area fixed effects and a vector of housing characteristics: year the structure was built, number of rooms, number of bedrooms, presence of a kitchen, presence of bathroom(s), and access to water. I then predict the metropolitan area-level rent using the fixed effects, and combine that with the share of consumer expenditure on rent (19.35% in 2000, which is lower than the typical 20%-30% for households in the US) to construct spatial price indices.

Figure 1.2a shows the distribution of the Laspeyres spatial price index, including or excluding housing, across the 46 metropolitan areas for which CPI price quotes are collected, using data from January 2000. Including housing costs, the 90th percentile metropolitan area is 21% more expensive than the median; the 10th percentile is 9% cheaper than the median. The large difference persists even after excluding housing costs from the price index: The 90th percentile metropolitan area is 14% more expensive than the median; the 10th percentile is 11% cheaper than the median. Excluding housing costs, the most expensive city, Mexicali—which is on the US-Mexico border—is 39% more expensive than the median city and 69% more expensive than the lowest-cost city (Tehuantepec, Oaxaca). Subfigure (b) shows the relationship between the Laspeyres price index excluding housing and average predicted housing cost at city level, in a log-log specification. Housing cost predicts non-housing prices in a statistically significant manner, with a coefficient of 0.15 and a t-stat of 3.1. Overall, the substantial dispersion in prices suggests that location-specific pricing is likely present in Mexico, which leaves room for location-specific shocks to affect local prices.

[5]Banco de México/INEGI takes a simple average of prices across all items within a product category, which is the lowest-level nest in the construction of the Consumer Price Index.

Product category-level price dispersion

Having established that large spatial variations in non-housing prices across Mexico are present, I provide further evidence that substantial price variation exists for items within the same product category. The ideal exercise would compare prices of the same item (i.e., with the same barcode/product specification) across locations. However, items included in the CPI data are often not exactly the same across space even for the same product and brand likely due to slight variations in the attributes of the most prominently displayed products across locations. For example, under the product category of "Cereals" in almost all cities Kellogg cereals are sampled, but the exact variety varies from "corn flakes, box of 500 grams" to "box of 600 grams" or "box of 350 grams." In some cities the description says "Kellogg's," but in others it says "Kellogg's Zucaritas," which suggests that they may be two different brands.

In light of this, I examine the average prices of items across locations **within the same product category**. Within each metropolitan area and product category, I compute the simple average of prices at a given point in time.[6] I then examine how these average prices across broader product groups at product category level differ across cities, by taking a weighted average of prices within a broad product group across product categories for each city, where the weights are the official expenditure weights attached to each of the 315 product categories. The results are similar if I take an unweighted average. Figure 1.3 shows the interquartile range, i.e., the difference between the 75th percentile and the 25th percentile, of prices across municipalities relative to the median (standardized to 1), using data from January 2000. Figure A.1 shows the same results for October 1993 and March 2018, the first and last months of my analysis sample; the results are similar. Panel A of Figure 1.3 shows that the 75th percentile city has 11% higher unprocessed food prices, 10% higher processed food prices, 25% higher non-food tradable goods prices, and 29% higher services prices than the city with median prices. This dispersion of services prices is expected, since services employ a much larger share of local labor as inputs, but the large dispersion of food and non-food tradable goods prices again shows that substantial geographic disparities in prices exist in Mexico. In Panel B, I only include tradables and drop services from my sample. Instead of grouping products into four broad groups, I break them down into eight finer groups based on the Classification of Individual Consumption According to Purpose (COICOP), a classification developed by the United Nations Statistics Division and widely adopted by agencies such as the United Nations, US Bureau of Labor Statistics, and European Commission.[7] I see that the tradable product groups that exhibit the largest spatial price variation are clothing and footwear, health products (medications, health equipment, etc.), and personal care goods (cosmetics, soaps, razors, etc.). Even though the dispersion of prices likely captures some differences in varieties across cities, these figures still provide evidence that prices are highly dispersed across Mexican cities,

[6] In calculating the Mexican CPI, Banco de México/INEGI takes a simple average of prices across all items within a product category.

[7] This classification of Mexican CPI data was first used by Gagnon (2009).

an environment in which location-specific shocks could have a location-specific impact on prices.

1.4 Empirical strategy

Identification: Shift-share instrumental variables (SSIV)

Problem. I set out to study the effects of local employment changes based on local labor demand shocks on price changes over five-year periods, $\log(p_{jt})-\log(p_{j,t-5})$. The independent variable of interest is $\frac{L_{jt}-L_{j,t-5}}{L_{j,t-5}}$, which is local employment changes over the initial stock of workers in municipality j between period t and $t-5$, and measures the percentage growth in local labor. By an accounting identity, the independent variable can also be written in a "shift-share" structure, combining pre-period employment shares in industry k with changes in employment in that industry:

$$\frac{L_{jt}-L_{j,t-5}}{L_{j,t-5}} = \sum_k \frac{L_{j,k,t-5}}{L_{j,t-5}} \left(\frac{L_{j,k,t}-L_{j,k,t-5}}{L_{j,k,t-5}} \right)$$

It can readily be seen from this equation that a significant source of endogeneity could arise if we were to estimate the effects on price changes using OLS. First, the expression in parentheses, $\frac{L_{j,k,t}-L_{j,k,t-5}}{L_{j,k,t-5}}$, combines both labor supply and labor demand responses. Places that experience a positive labor demand shock may also experience a positive or negative labor supply shock arising from demographic, human capital, or migration changes, and that could affect prices by changing the aggregate demand for goods and services. Second, changes in local employment in location j and sector k can be driven by location-specific cost shocks, such as a decrease in the cost of inputs due to the discovery of natural resources. Such shocks would plausibly be correlated with variation in prices in location j if, for example, they lead to lower marginal costs of locally produced goods. Finally, local infrastructure development can both attract firms and lower trade costs, which would create biases in estimates of the effects of local labor demand shocks on prices.

SSIV setup. To address these endogeneity concerns, I introduce a shift-share instrumental variable (SSIV) that leverages national changes in labor demand to isolate quasi-exogeneous local labor demand shocks.[8] The instrument was first popularized by Bartik (1991) and has been widely used in empirical labor and international economics (Blanchard and Katz, 1992; Card, 2001; Autor, Dorn, and Hanson, 2013; Diamond, 2016). To start, I

[8] We could start by considering other sources of policy variations to identify a valid instrument for labor demand. In general, however, natural experiments that affect local labor demand on a national scale are hard to come by. Absent large-scale policy changes, such shocks are extremely difficult to find. Some studies use quasi-random manufacturing plant placements (Greenstone, Hornbeck, and Moretti, 2010; Abebe, McMillan, and Serafinelli, 2022); however, even if such quasi-random placement exists, the arrival of one specific plant likely has very limited effect on employment in large municipalities.

can write the SSIV as follows:

$$\text{SSIV}_{jt,t-5} = \sum_k \frac{L_{j,k,t-5}}{L_{j,t-5}} \left(\text{Predicted Labor Demand Growth}_{j,k}\right)$$

where I use employment changes in municipalities other than j to predict municipality j's labor demand growth:

$$\text{SSIV}_{jt,t-5} = \sum_k \frac{L_{j,k,t-5}}{L_{j,t-5}} \left(\frac{L_{-j,k,t} - L_{-j,k,t-5}}{L_{-j,k,t-5}}\right)$$

In this expression, $L_{-j,k,t}$ denotes employment in industry k in year t in all municipalities except j (a "leave-one-out" specification). For industries k, I use 3-digit NAICS industry codes. There are 89 3-digit industries across Mexico in my data.[9]

In my setting, the instrument leverages changes in national labor demand at industry level as a source of exogenous variation, which could stem from either industry-specific productivity changes in Mexico or global demand shocks for products produced in Mexico. Across my study period, the observed variation stems from trade liberalization and the expansion of manufacturing and services industries across Mexico. Identification follows from a shock-level exogeneity condition, first formalized by Borusyak, Hull, and Jaravel (2022), which transforms the location-level regression to an equivalent shock-level one in which shocks are weighted by locations' exposure to them. Under this assumption, industries that experience a rise in labor demand should not face systematically different unobservables that affect both local income and local prices in locations with the highest share of those industries. Under the assumption that identification comes from the exogeneity of shocks, exposure shares are allowed to be endogenous. In Section 1.4, I discuss in detail the identifying assumptions and assess the validity of those assumptions.

The research question of interest focuses on how changes in local nominal income driven by a local labor demand shock affect prices. In light of that, I run both (a) reduced-form regressions to study the direct effects of changes in local labor demand, and (b) IV regressions, instrumenting local income changes with the shift-share instrument, to study the price effects of changes in nominal income driven by labor demand shocks. The regressions are implemented in first differences:

$$\log\left(\frac{P_{igjt}}{P_{igjt-5}}\right) = \alpha_{gt} + \beta \text{SSIV}_{jt} + \epsilon_{igjt} \qquad \text{(Reduced form)} \tag{1.2}$$

$$\log\left(\frac{P_{igjt}}{P_{igjt-5}}\right) = \alpha_{gt} + \beta \log\left(\frac{w_{jt}}{w_{j,t-5}}\right) + \epsilon_{igjt} \qquad \text{(IV)} \tag{1.3}$$

where j denotes municipalities, g denotes product categories, i denotes items (barcode-by-store combination) and t denotes census year. $\log\left(\frac{w_{jt}}{w_{j,t-5}}\right)$ is the log changes of wage bill per

[9] My preferred specification uses percentage changes to construct changes in industry-level employment, but the results are also robust to using log changes.

employed worker in municipality j, instrumented with the Bartik labor demand shock. Since the regression is implemented in first differences, it already absorbs the barcode-by-store (item)- level fixed effect; an observation is an item present in both year t and year $t-5$. α_{gt} controls for unobserved product category-by-time trends such that the regression compares the inflation of products within each one of the 315 product categories, e.g., inflation of automobiles in Guadalajara vs Veracruz.

To rule out changes in product characteristics that could affect the price trajectory of items, I only include in my main analysis products that were not substituted during the five-year period, and remove observations that experienced a change in product characteristics (e.g., changing from Kellogg's Corn Flakes to Zucaritas for the same item series in the CPI data). To ensure that my sample reflects a representative basket of consumer goods and services, I follow Atkin, Faber, and Gonzalez-Navarro (2018) and construct weights such that the total weight of observations in each product category matches the consumption weights in the official CPI manual, which I use for all item-level regressions. The coefficient β thus captures the effect of a municipality-level labor demand shock on a representative item in the consumer's consumption basket.

Equations (2) and (3) are performed at the item level, which allows me to compare items within product categories. A more traditional approach for SSIV regressions would be to use the municipality-level price index as the dependent variable. By collapsing price changes across items into an index, this approach does not offer clean, apples-to-apples comparison (e.g., a car in municipality A is being compared to all kinds of products in municipality B), but it does offer a more intuitive interpretation. A municipality-level regression also offers the advantage of being able to construct exposure-robust standard errors a la Adão, Kolesár, and Morales (2019), which takes into account correlation across national shocks and provides another robustness check of the validity of the instrument. As a result, I implement municipality-level regressions with the Laspeyres price index $P_j^{t,t-5}$, defined as

$$P_j^{t,t-5} = \sum_{i \in I} \frac{p_{i,t}}{p_{i,t-5}} s_{i,t-5} \tag{1.4}$$

for municipality j, where i denotes product categories and $s_i, t-5$ is the expenditure share on category i (which is always constructed from consumption survey data in the base period; thus, the -5 subscript). To construct P_i from product category-level prices, I follow the approach of Banco de México/INEGI and take a simple average of prices across items in the lowest nest, across varieties and sampling locations. When I regress at the municipality level I weight each cell by the number of items, i.e., price quotes.

SSIV strongly predicts local employment growth. Figure 1.4 shows that the shift-share instrument correlates strongly with local employment growth. The left subfigure shows the binned scatter plot at the municipality-year level between log employment growth and the shift-share instrument after absorbing year fixed effects. The regression coefficient is 1.06 with a t-statistic of 9.69. On average, across periods and municipalities in my data, local employment increases by 20% over a five-year period. This corresponds to an average annual employment growth rate of $(1 + 20\%)^{\frac{1}{5}} - 1 = 3.71\%$.[10] This is higher than the 1.5%-2% population growth rate observed in Mexico over the study period, suggesting that employment growth in urban areas in my sample consists of both population growth and migration into cities. The right subfigure shows the binned scatter plot between changes in the employment-to-population ratio and the shift-share instrument, which can account for changes in local labor due to migration responses and serves as another proxy for the health of the local labor market. The regression coefficient is 0.23 with a t-statistic of 6.1.

Identifying assumptions

Quasi-random Assignment of Shocks. For the shift-share IV to causally identify the effect of local labor demand shocks on prices, two assumptions must be satisfied: First, shocks should be as good as randomly assigned, i.e., changes in national sector-level employment rates should be uncorrelated with potential confounds that affect prices, and second, the shocks are large in number and sufficiently mutually uncorrelated, which would be standard assumptions to establish the consistency of the IV estimator (Borusyak, Hull, and Jaravel, 2022). Consistency follows from the 89 distinct 3-digit sectors in my analysis; these follow the Mexican SCIAN classification, which in almost all cases are consistent with sectors in the North American Industry Classification System (NAICS). To test that the shocks are sufficiently mutually uncorrelated, I construct the intra-class correlation coefficients (ICC) at industry level. The value is 0.188, which reveals moderate clustering of shock residuals, but the degree is similar to the ICC of shocks in other shift-share studies such as Autor, Dorn, and Hanson (2013) (0.169).[11]

Because my estimation is in first differences, the first assumption—quasi-random assignment of shocks—is similar to a parallel trends-style assumption at shock (industry) level. I test this assumption in three ways. First, I regress price changes within one period on the value of the next period's shift-share IV, which is in the same spirit as the pre-trends test commonly employed in difference-in-differences specifications. Second, I conduct industry-level balance tests by regressing pre-exposure (i.e., baseline) outcomes on the shift-share IV. Third, I conduct additional location-level balance tests with variables that measure demographics and labor market conditions. The latter two tests aim to provide additional evidence in support of the orthogonality of shocks, though they are not required for identification because the causal model is in changes instead of levels.

First, to test for pre-trends, I regress price changes on the future values of the shift-share instrument. Specifically, I regress 1993-1998 changes in prices on the value of the shift-share IV constructed from the 1998-2003 difference, and regress 2003-2008 changes in prices on the 2008-2013 shift-share IV. If shocks are randomly assigned in changes, the future-period shift-share IV should not affect past-period price changes. If there is a correlation, we might worry about pre-trends: Locations with large increases in predicted local labor demand

[10] The 10th percentile is 10%, the median is 20%, and the 90th percentile is 29%.

[11] I compute the ICC using a hierarchical random effects model following Borusyak, Hull, and Jaravel (2022).

may already be on a different price growth trajectory from locations with small increases in predicted labor demand. Panel A of Table 1.1 reports coefficients and standard errors from the pre-trends tests. Reassuringly, the future period's shift-share IV has no statistically significant effect on price growth, for both goods and services items.

Following Borusyak, Hull, and Jaravel (2022), I also conduct balance tests on pre-exposure levels, not simply changes, of industry- and location-level variables. Even though the quasi-random assignment of shocks only needs to hold in changes, not levels, the results can help assuage the concern that industry-level shocks may be correlated with other industry-level variables that systematically affect the price evolution of products. I do this by first regressing industry-level potential confounding variables directly on industry-level shocks. I choose the industry-level variables that are available in the Economic Census and might be correlated with the shift-share instrument: start-of-period log labor income, log employment, and log value added. I also regress each sector's average employment share on the instrument to test whether the instrument is affecting places with a higher or lower share of local employment. In the regressions, I include period fixed effects and weight each industry-period by average industry employment shares in that period, as required for shock-level SSIV regressions. I also test for balance on location-level demographics variables: share of adults with secondary school education, share of adults with university education, share of adults who speak an indigenous language, share of workers who are women, share of workers in manufacturing, and share of workers in services.[12] These variables broadly reflect the composition and characteristics of a location's workforce. All variables are constructed from decennial population census data, which took place three years prior to the beginning of each of my analysis periods (e.g., the population census took place in 1990 and the first analysis period was 1993-1998). To obtain exposure-robust standard errors that account for the correlation across observations with similar exposure shares (Adão, Kolesár, and Morales, 2019), I implement these regressions at shock level.

Panel B of Table 1.1 reports the results of industry-level balance tests. All pre-period industry-level variables, which on a broad level reflect the employment and health of industries, are balanced. Panel C of Table 1.1 reports the results of location-level balance tests. I find that the shift-share instrument is not correlated with start-of-period education attainment or share of employment in services, but is significantly and positively correlated with the start-of-period share of female workers among workers and share of employment in manufacturing, and negatively correlated with the start-of-period share of adults who speak an indigenous language. The positive correlation between the instrument and lagged manufacturing employment share is unsurprising: Some of the top industries that experienced employment growth are in manufacturing—specifically, transportation equipment, food, and apparel. It is unsurprising that places with higher employment growth rates tend to be those with a high baseline share of manufacturing. Even though identification does not require balance on levels—the causal model is in first differences—the observed imbalance on

[12]Since my sample consists of urban municipalities, agriculture accounts for a negligible share of employment. The mean is 3.8% and the median is 1.8% across municipality-periods.

pre-period location-level variables may generate concerns that predicted employment growth is higher in places with high baseline manufacturing activity or female labor participation, and that these places also have unobserved trends in common. In my analysis, I evaluate the sensitivity of the results by controlling for all start-of-period location-level variables plus start-of-period logged wages.[13]

Exclusion Restriction: Causal Channel Only Runs Through Income. Also, for the shift-share instrument to identify the causal effect of local income changes on local price changes, the standard IV exclusion restriction needs to be satisfied. This requires the assumption that predicted changes in local labor demand do not affect prices through any channel other than through changing local income—i.e., by changing the demand for goods and services. I argue that this is a reasonable assumption for several reasons. First, the leave-one-out structure of the instrument ensures that local productivity effects do not drive the results. Since construction of the instrument uses national leave-one-out changes in employment, productivity changes specific to one municipality (e.g., municipality-level subsidies or road construction, the arrival of a multinational plant) do not enter into the instrument. Nevertheless, we might still worry about correlation between local employment shares and national employment changes, which would imply that a given productivity shock in a sector would induce larger or smaller decreases in the marginal cost of goods produced by that sector in areas with higher employment in that sector. To address this, in one specification I construct my shift-share instrument to be product category-specific: For each one of the 315 product categories, I leave out the 3-digit sector that most closely contributes to the production of those goods. For example, the shift-share instrument for all clothing products would exclude employment in NAICS 315 (apparel manufacturing). I present the results in Section 1.5.

An additional concern is that labor demand shocks could lead to trade cost decreases if, for example, local governments invest more in public infrastructure in response to the opening of new manufacturing plants. Such public investments could decrease local inflation by allowing goods to be transported more easily and quickly. In a robustness exercise in Section 1.5, I control for municipality-level public investment and show that the results are not driven by such investments.

The SSIV capture industry-wide labor demand shocks. Even though the leave-one-out structure of the SSIV excludes local labor supply shocks, there remains a concern that instead of being driven by industry-level labor demand changes, employment changes in the shift-share instrument may capture national, industry-wide labor supply changes. This would lead to different interpretations of the empirical results. While I cannot conclusively rule out the presence of such labor supply shocks, the fact that I observe a positive first

[13] Both the instrument-confounder and outocme-confounder relationships may be nonlinear. To address nonlinearities in the confounding relationship, one typically uses double/debiased machine learning (DML) (Chernozhukov et al., 2016). However, classic DML methods do not allow for fixed effects. Clarke and Polselli (2023) extends DML by separating the impact of confounding into a component with nonlinear effects and a component with fixed effects. I have not found an available Python package, but once it is available I will implement their method to nonlinearly control for potential confounders.

stage at the sector level (discussed in the next section) shows that wages respond positively to predicted employment increases from the SSIV, suggesting that the effect is likely primarily driven by an outward shift of the labor demand curve instead of the labor supply curve. I also examine the 3-digit industries that experienced the largest employment increases in percentage terms during each one of my three analysis periods. Appendix Table A.1 displays the industries. The industries listed, such as administrative and support services, transportation equipment manufacturing, apparel manufacturing, and food services, belong to manufacturing and services industries that benefited from Mexico's integration into NAFTA and structural transformation. The former increased export demand for manufacturing goods—for instance, the Mexican auto industry was a prominent success story that grew more than threefold in output from 1994 to 2016—and the latter increased demand for services (Klier and Rubenstein, 2017). These results should provide further confidence that industry-level labor demand shocks were the main drivers of sectoral employment growth observed in the data.

Potential SUTVA Violation. An important but commonly hidden assumption in empirical research is the stable unit treatment value assumption (SUTVA): Potential outcomes for a given observation should respond only to its own treatment status. In my setting, SUTVA stipulates that there are no spillovers between locations exposed to different doses of labor demand shocks. This assumption would be violated if, for example, retail prices are set according not only to municipality-level but also city- or state-level market demand. In that case, a municipality's retail price would be affected by neighboring municipalities that are experiencing positive labor demand shocks. To the extent that this occurs in my setting, the violation would imply that the estimated effects are a conservative lower bound: Spillover effects would bias the estimates toward zero, because non-affected municipalities experience a price effect in the same direction as affected neighboring municipalities. In ongoing work, I am developing a robustness exercise that only includes municipalities that are not geographically connected to each other.

Robustness to Heterogeneous Treatment Effects. The causal model assumes homogeneity in treatment effects by specifying one β that does not depend on municipality j or product i. However, the true effect of a local labor demand shock on prices may differ across locations or products in ways that are not ex ante obvious to researchers. In their appendix, Borusyak, Hull, and Jaravel (2022) show that the shift-share instrument is robust to heterogeneous treatment effects. SSIV estimates a convex weighted average of heterogeneous treatment effects as long as a first-stage monotonicity condition is satisfied. In my setting, the monotonicity condition requires that predicted local labor demand growth weakly increases local labor income. This would be satisfied given the typical assumption that the aggregate labor supply curve in local labor markets is upward-sloping.[14]

[14] Also, to show the strength of my instrument, I visualize the reduced-form and the first-stage relationships in Appendix Figure A.2. The figures present the binned scatter plots at the 3-digit industry level between the sector-level shock—i.e., labor demand changes—and (a) changes in log prices and (b) changes in log labor income, following the shock-level transformation in Borusyak, Hull, and Jaravel (2022).

1.5 Results

Main results

Table 1.2 presents the main results from both the reduced-form and IV regressions, where the instrumented endogenous variable is the change in log labor income per worker. I use data from three five-year differences: 1993-1998, 2003-2008, and 2003-2018. In the reduced-form regression, the main explanatory variable is the shift-share instrument, which predicts changes in local labor demand. In the IV regression, I instrument changes in log wage bill per worker with the shift-share instrument. The explanatory variables are at municipality level for the last two differences, and at metropolitan area level for the first difference because data constraints made it impossible to identify the municipality of the price quote before 2002. All standard errors are clustered at metropolitan area level. All regressions include product category by year fixed effects to control for unobserved trends in the price evolution of items across different product categories, and are weighted such that the sum of weights within a product category matches the official weights used for the CPI calculation. I also adjust the weights so that each of my three five-year periods receive equal weight in the regression.

Column 1 shows that across all product categories—goods, energy, and services, the predicted labor demand shock has no effect on prices. However, the overall null result masks substantial heterogeneity. Columns 2 to 4 show that a 1 percentage point increase in the growth rate of predicted local labor leads to a 0.166 percentage point increase in energy prices inflation and a 0.072 percentage point increase in services price inflation (i.e., the growth rate of prices), but a 0.057 percentage point *decrease* in goods price inflation.[15] Columns 5 and 6 show that the negative result for goods inflation and the positive results for on services inflation are robust to adding control variables that account for start-of-period demographics and local labor market controls. In Column 7, I address the concern that cost-reducing shocks to specific sectors may lower prices more in places with higher employment in those sectors if, for example, certain technologies disproportionately benefited places with high baseline employment in those sectors. As discussed in Section 4.2, for each product category, I remove the 3-digit sector that produces those goods in constructing the shift-share instrument. Since this pushes the analysis into an "incomplete shares" environment where the sum of shares across sectors does not add up to one, I follow best practices for those cases and control for the sum of shares in the regression. The coefficient is still negative, statistically significant, and close to that in Column 5 (-0.0865 vs -0.0991). I also use a version of the leave-out shock in constructing the instrument for examining effects on services, and the effect remains positive and statistically significant.[16]

[15] The log transformation of the dependent variable, $\frac{P_t}{P_{t-5}}$, allows us to interpret the coefficient as the effect on the growth rate of prices: When $\frac{P_t}{P_{t-5}}$ is close to 1, $\frac{P_t}{P_{t-5}} - 1 \approx \log(P_t) - \log(P_{t-5})$.

[16] All regressions cluster standard errors at the metropolitan area level, which accounts for both within-metropolitan area correlation across municipalities and across-time correlation within the same metropolitan

The bottom part of Table 1.2 reports IV regression results, which enables us link changes in local income with changes in prices directly. From Columns 7 and 8, we can see that a 1 percentage point increase in the growth rate of a location's labor income per worker leads to a 0.171 percentage point decrease in the rate of the price growth of a representative tradable item at that location, and a 0.169 percentage point (statistically insignificant) increase in the rate of the price growth of a representative services item at that location. Intuitively, a positive local labor demand shock pushes up wages and rents. Since labor and rent constitute a larger share of input costs for services production than goods production, especially at the customer-facing stage (i.e., retail), the effect on services prices is expected to be positive and larger than the effect on goods prices. However, the effect on tradable prices is perhaps surprising: As local income increases, we would expect local demand for tradable goods to increase, which in turn would drive up prices. One thing to note is that the negative coefficient does not mean that higher income-growth places experienced deflation. Over the five-year periods, prices still increased in all municipalities, albeit more slowly in places experiencing income growth. To dive deeper into these results, in the following sections I first conduct various robustness checks to verify the validity of the results. I then explore the mechanisms that could drive the relative decrease in prices I found.

The regression in Table 1.2 is at the item (i.e., barcode-by-store combination) level and accounts for differences in the price trajectories across different product groups. Another way of analyzing the effects is to collapse the data to municipality level so that we can examine the effect of local labor demand shocks on a local price index—a measure of changes in consumers' purchasing power. Table 1.3 reports the results where the dependent variable is the local Laspeyres price index, defined in Section 4.1. The reduced-form regression shows that a 1 percentage point increase in the growth rate of predicted local labor demand leads to a 0.05 percentage point decrease in the Laspeyres price index for goods items. The IV regression shows that a 1 percentage point increase in log wages per worker leads to a 0.12 percentage point decrease in the Laspeyres price index. The magnitudes of the coefficients are lower than those in Table 1.2. As expected, the two sets of coefficients differ since the Laspeyres index does not control for product category-specific trends, but it is nonetheless reassuring that the results are similar and point in the same direction: An increase in local labor demand decreases local goods inflation and increases local services inflation, though the effect on services inflation is often noisily estimated.

The first stage regressing log wage changes on local labor demand changes also traces out the inverse labor supply curve. The first-stage coefficient, which is the ratio of the reduced-form coefficient to the IV coefficient, corresponds to the inverse of the labor supply elasticity and is about 0.4 to 0.5. The spatial labor supply elasticity I estimate here—the inverse of my first-stage coefficient—is about 2 to 2.5, which is similar to estimates commonly found

area. Alternatively, if I cluster standard errors at metropolitan area level for 1993-1998 and the municipality level for the other two periods, the p-value on the reduced-form coefficient in a specification without controls (column 3 of Table 1.2) changes from 0.079 to 0.086, and the p-value in a specification with controls (column 5 of Table 1.2) changes from 0.025 to 0.028. IV coefficients also remain statistically significant (p-value of 0.003 vs 0.002 with alternative clustering).

in the literature that are around 2 (Fajgelbaum, Morales, et al., 2019).

Heterogeneity across product groups

Which product groups are most responsive to demand shocks? Examining the heterogeneity of the effect of local labor demand shocks on price changes helps us achieve two objectives. First, it can provide an intuitive check on the validity of the main results. Certain product groups, such as packaged snacks and beverages, exhibit smaller price variation across space since they are often shipped from the same manufacturing plants in the country (or imported) and carry pricing suggestions by the manufacturer. Other products, such as clothing, cars, and appliances, exhibit larger price variation across space because retailers price them according to location-specific consumer demand, which can differ widely for the same item. This may also imply that price growth in these categories can vary more freely with local demand conditions. Second, heterogeneity analysis can inform us about the distributional impacts of price effects from local labor demand shocks. Slower price growth concentrated in, for example, food products will likely yield higher gains for low-income consumers in fast-growing cities than concentrated in, say, cars and electronics.

Figure 1.5 shows product group-specific coefficients among tradable goods in a reduced-form regression of log price changes on the shift-share instrument, where the instrument is interacted with product group dummies. Product groups are constructed according to the Classification of Individual Consumption According to Purpose (COICOP), a classification developed by the United Nations Statistics Division.[17] We can see that the effects are strongest on unprocessed food, clothing and footwear, transport equipment (cars, bicycles, etc.), recreation, sports and culture goods (radios, video games, sports equipment, etc.), and personal care goods. These groups of products are likely sold by retailers with more discretion in tailoring prices, and thus markups, to local demand conditions—unlike, for example, processed food and alcohol, which are primarily produced by large manufacturers and are more likely to feature more uniform pricing across locations. Product categories with the largest effects also closely correspond to those that exhibit the largest variation in prices across space, as illustrated in Figure 1.3(b).

Robustness of results

In the preceding sections, I showed that an increase in local labor demand leads to a small but statistically significant decrease in local tradables prices, and that the results are not driven by a direct productivity effect that lowers the local marginal costs of production. Here I test for the robustness of the results and consider alternative interpretations.

Appendix Table A.2 reports the results of a specification that controls for region-by-period fixed effects. This amounts to controlling for region-by-period-specific trends, since

[17] COICOP only has one group for food items, which I break it down into processed and unprocessed food following the methodology used by Gagnon (2009) with the same Mexican CPI data.

my regression is in first differences. Regions are defined according to the common definition employed in INEGI surveys: Northeast, Northwest, Southwest, Southeast, West, East, and Central South. Results for tradables are quantitatively close to results from the main specification although the results for services become not statistically significant. This suggests that region-specific labor market or price trends are unlikely to drive the main results, consistent with the identifying assumptions of the shift-share instrument.

One additional concern is that the IV endogenous variable, changes in log wage per worker, may not be robust to compositional changes in the labor force. An example of this phenomenon is that at the peak of the COVID-19 pandemic in 2020, worker earnings increased in the US. However, this spike in wage growth primarily reflects the departure of low-wage workers who had been laid off (Crust, Daly, and Hobijn, 2020). To address this concern, in Appendix Table A.3 I construct the IV endogenous variable to be changes in log wage per adult rather than per worker, where the municipality-level adult population is obtained from a linear interpolation of decennial census figures. This is because the Economic Census, from which my variables are constructed, takes place three years after or two years before each decennial population census.[18] The point estimates are very close to those from the main specification, although the IV F-stats are slightly reduced. This is likely due to measurement error in the "log wage per adult" variable, since it involves interpolation which decreases the power of the first stage.

Another concern is that places that experience an increase in labor demand also experience a decrease in trade costs, which then lowers the prices of tradables. Recall from the discussion in Section 4 that if the decrease in trade costs is specific to one municipality, it is unlikely to be an issue since the shift-share instrument uses a leave-one-out approach. However, we might still be concerned about industry-level correlations between labor demand and trade costs: for example, if the petroleum industry experienced a boom and it led to both higher employment and increased road construction, then the causal channels might include trade costs in addition to income effects. To test this story, I control for log changes in public spending from the State and Municipal Public Finance Statistics dataset (*La Estadística de Finanzas Públicas Estatales y Municipales, EFIPEM*), which reports annual municipality-level expenditure on public works and projects, social services, general services, etc.[19] It has the advantage of including spending at all levels: federal, state, and municipal. Absent data that directly quantify trade costs across municipalities in Mexico, public spending, especially spending on public works, can be a good proxy for trade costs because it accounts for both infrastructure construction (highways, railways, waterways) and infrastructure maintenance, both of which could affect the ease of transporting goods across locations. In Table 1.4, I control for the 5-year log difference of both the public investment expenditure, which includes public works and projects, and total municipality-level expenditure. I focus on tradable goods only since services are produced locally. The coeffi-

[18] An adult is defined as anyone aged 18-65.

[19] In the dataset, road construction is broadly defined and includes roads, highways, railroads, and waterways.

cients remain stable: For example, controlling for changes in public expenditure changes the reduced-form coefficient from -0.099 to -0.098.

An additional concern is that the results could be driven by growth in northern Mexican states close to the US border, which disproportionately benefited from NAFTA. During the first 10 years following the implementation of NAFTA, from 1994 to 2003, the six northern border states achieved an average annual GDP growth rate of 3.97%, which is higher than the 2.76% national growth rate (Randall, 2015). The implementation of NAFTA could lead to faster manufacturing growth rates in those states, which tend to have high manufacturing shares of employment to begin with. They may have also disproportionately benefited from access to imports from the US, which in turn could slow goods inflation. To address this concern, in Appendix Table A.4 I perform my main analysis on a sample excluding all municipalities in the six border states (Baja California, Chihuahua, Coahuila, Nuevo León, Sonora, and Tamaulipas). Results remain statistically significant and quantitatively similar to those for the full sample.

I also examine whether my results were driven by a particular adverse event: the Mexican peso crisis. There may be a concern that the peso crisis, which began in December 1994 and led to an immediate inflation spike in the first quarter of 1995, may have had a smaller effect on locations that experienced a positive nominal income shock. I address this concern in Table A.6 by testing whether price changes during the peso crisis were correlated with the shift-share instrument. I construct the price change of items from averages in the third quarter of 1994 to the second quarter of 1995, and regress that price change on changes in predicted local labor demand between 1993 and 1998 as a placebo check. I find a positive and not statistically significant coefficient in both the reduced-form and IV specifications, which indicates that the negative results for goods inflation in the main analysis are unlikely to be the consequences of the immediate inflation spike at the onset of the peso crisis.

Mechanisms leading to higher services inflation

An increase in market size due to the positive labor demand shock leads to lower tradable inflation among continuing varieties, but a positive, albeit borderline statistically significant, effect on services inflation. I first seek to explain the positive result for services, which is likely due to three reasons. First and foremost, services are much more exposed to increases in local wages because a larger share of costs is in local labor compared with goods production, especially goods that are not produced locally. Even if the same forces that lead to the entry of new goods varieties are present in the services sector, such forces are likely offset by the wage increases that compel firms to raise prices. Housing prices also likely increase, which further increases costs for local services production.[20] Second, even though services price quotes are collected for the same service at the same establishment, there are likely unobserved quality changes that accompany local labor demand shocks (Moulton et al., 1997;

[20]Because housing costs are noisily measured in Mexican CPI data, I do not include them in my main analysis. Nevertheless, I find that a 1 percentage point increase in the growth rate of local labor demand leads to 0.2 percentage point higher housing inflation.

Nordhaus, 1998). Whereas tradable price series measure the prices of the same product, service items focus on specific kinds of services that can be broader in definition, are harder to measure, and differ across time. For example, the CPI microdata include a price quote for "car repair and maintenance: changing oil and filter." This observation measures the price at the same establishment, but does not account for the length or quality of the service. Establishments in places experiencing a positive shock may be more likely to upgrade their service quality due to changes in consumer taste, which can increase prices. Lastly, services products are likely less substitutable for each other compared with tradables. Consumer demand for services is likely more inelastic, especially for services offered by sellers with whom consumers enjoy a personal relationship (e.g., barbers, bakers, doctors, travel agents) than for tradable goods. This can lead to smaller decreases in the price index, even if local labor demand shocks lead to the entry of new services establishments, since consumers tend to stick with their existing services establishments.

1.6 Mechanisms leading to lower tradable goods inflation

It remains to be explained why local labor demand shocks lower local inflation among tradable goods. In this section, I aim to uncover the potential mechanisms and the ways in which I would proceed to test each of those. Price equals markup times marginal cost, so the effect must come from a reduction in markups, marginal cost, or a combination of both. Overall, the evidence suggests that a reduction in markups driven by pro-competitive pressure from firms introducing new varieties seems to be the most plausible story.

Marginal cost

One hypothesis is that the reduction in price following a local income shock could be driven by a reduction in the marginal cost. The previous section shows that a contemporaneous trade cost decrease is unlikely to be the source behind the relative decrease in prices. In this subsection I consider an alternative channel that could drive down the marginal cost following a positive local labor demand shock: external economies of scale (EES), whereby at the industry level the private marginal costs of production are higher than social ones, and the industry-level supply curve is actually downward-sloping (Bartelme et al., 2019). Given that my effects are local to a municipality, in my setting such external economies of scale, if they exist, are more likely to be present at the wholesale/retail level instead of at the manufacturing level, since tradable products are typically manufactured outside the municipality in which they are sold. Earlier literature has argued that external economies of scale may exist in the retail sector (Ofer, 1973; Kato, 2012), and could stem from sharing physical infrastructure (mall space, cargo offloading facilities), increased productivity in sourcing, or productivity spillovers across workers.

If this is the case, it would in turn suggest that income effects can drive productivity effects by driving up scale, which are of key policy interest. Economists and policymakers have long been interested in productivity gains and spillovers to firms in developing countries and have studied many channels, such as exporting (Atkin, Khandelwal, and Osman, 2017), joining multinational supply chains (Alfaro-Ureña, Manelici, and Vasquez, 2022), observing and imitating foreign firms (Abebe, McMillan, and Serafinelli, 2022), and improving management practices (Bloom et al., 2013). A finding that income shocks could lead to productivity gains would point to another channel through which productivity improvement in developing-country firms can occur and emphasize that a comprehensive welfare analysis of local income shocks should take into consideration the effect on the price index through the productivity channel. The results could also highlight additional productivity benefits of cash transfer and poverty alleviation programs.

A necessary condition for external economies of scale is that wholesale and/or retail employment should increase following an increase in predicted local labor demand as the sector expands. I examine this hypothesis in Panel A of Table 1.6. I regress municipality-level changes in wholesale and retail employment on my shift-share instrument, which captures changes in predicted local labor demand. I also break down employment into subcategories of retail employment to examine whether specific retail sectors (such as modern retail: department stores and convenience stores) experienced expansion. I find that the shift-share instrument has no effect on overall wholesale or retail employment. It leads to a decrease in overall retail employment (albeit not statistically significant), a decrease in retail employment in food and beverages, and an increase in employment in department stores, which were a relatively new and modern kind of retail establishment in Mexico during that time. The negative effect on overall retail employment—and specifically on employment in food and beverage retail, which primarily consists of mom-and-pop stores and market stalls—may point to a reallocation of labor away from unproductive segments of the retail sector toward modern retail as a result of local labor demand shocks. These results might suggest external economies of scale in the specific subsector of department stores. However, department stores only constitute between 2.7 to 3.7 percent of retail employment in my sample of urban municipalities across census years between 1993 and 2018. In Appendix Table A.5, I exclude employment in modern retail—supermarkets, mini-supers, and department stores—from my shift-share instrument to further test this hypothesis. The results are quantitatively similar. Taken together, the expansion in department stores only increases retail supply by a small amount, and the expansion alone seems unlikely to generate a significant price effect even under a model of sector-specific external economies of scale.

As I will discuss in the next subsection, the expansion in employment in department stores is accompanied by an increase in the number of department store establishments. In light of other pieces of evidence that I will discuss, I interpret the increase in the number of establishments of and employment in department stores as evidence of pro-competitive pressure, rather than external economies of scale. Overall, it does not seem that external economies of scale are driving a fall in the marginal cost that could drive down consumer prices.

Markups

Could a relative decrease in markups drive the result? Recall that my analysis sample consists of products with the same brand and specification sold by the same store over a five-year period, so these represent existing varieties that were already available to consumers rather than newly introduced varieties. The prices of these existing varieties could fall (or increase less, since overall households still experienced inflation) in markets that experienced positive labor demand shocks, and therefore income growth, if firms respond to local income growth by skewing product innovation and introductions toward those markets. An increase in the number of new varieties then leads to a higher number of substitutes for existing varieties, which exerts competitive pressure and drives down the prices of those older varieties.

Jaravel (2019) illustrates a story of this flavor in their study of inflation inequality in the US: Richer households experience lower product inflation because firms introduce more new products, which drives down markups of continuing products due to competitive pressure.[21] An increase in demand, driven by firms' introduction of more product innovations to growing market segments, leads to a substantial fall in prices and increases in product varieties, because increased competitive pressure from new products pushes markups down. It is possible that in my setting a mechanism similar to Jaravel's is taking place but in a spatial dimension. Trade liberalization that began in the 1980s has led to uneven growth and economic disparities within Mexico (Aroca, Bosch, and Maloney, 2005); as a result, it is entirely possible that new varieties, either from imports or domestic production that targets higher-income consumers, are being introduced more frequently in higher-labor-demand, higher-growth locations.

To test the hypothesis that an increase in product innovation that targets higher-growth locations drives down the inflation of existing products, ideally we would examine changes in markups directly. However, data on markups or marginal costs—from which one can calculate the markup—for consumer products are typically hard to come by; it is especially difficult to find marginal cost data in my context, as even the coarser Producer Price Index (PPI) data at product category level are not available prior to 2011. Absent data on marginal costs, I present several pieces of evidence that suggest there is likely competitive pressure from newly introduced varieties that drives down inflation among existing varieties. First, I examine the price changes of products that experienced a substitution. These are changes in item characteristics, such as model, brand, weight, and presentation, that authorities made because they deemed newer varieties to be more representative of the urban consumption basket. For example, in September 2002 in Monterrey a previous item "Pineapple slices in syrup, can of 800 grams, Herdez brand" was replaced by "Pineapple slices in syrup, can of 800 grams, La Costeña brand" at the same store, because the Herdez canned pineapple was deemed more representative of the consumption basket. In the same month and same city,

[21]Using Nielsen scanner data and consumer expenditure survey data, Jaravel (2019) finds that annual inflation for retail products was lower for higher-income households, and that much of the effect comes from lower price growth among existing varieties.

"Pine wood bed, 5 pieces, brown, Savanna Brand" was replaced by "Poplar wood bed, 5 pieces, brown, Verona Brand" at the same store because the latter was introduced as a new model. As these examples illustrate, products that undergo substitution are still the same kind of product but with different characteristics. If the effects operate through decreasing markups on existing varieties, we would expect a less negative or positive effect on products that experience substitution, since those products are therefore less likely to be subject to competitive pressure from new varieties.

Table 1.5 reports results of my main specification on this sample of products. The coefficients are positive across all specifications, for both goods and services, though not statistically significant. These results suggest that an increase in predicted local labor demand leads to a possible *increase* in the price of items that experienced a change in characteristics, contrary to the results in Table 1.2 which reports a decrease in the price of continuing varieties whose product characteristics were constant. Relative to low-income-growth locations, in high-income-growth locations products in which there is more innovation follow a different price trajectory: Products for which newer varieties become available are becoming relatively more expensive, not cheaper, in high-income-growth areas. Comparison of the results between Table 1.5 and Table 1.2 suggests that existing varieties follow a different relative price trajectory in high- vs low-income-growth areas, which suggests the presence of market forces generated by local labor demand shocks that specifically affect existing varieties, but not new varieties.

I also regress the effect of the shift-share instrument on the frequency of product substitution in the CPI data, which serves as a proxy for the degree of product innovation. Although idiosyncratic reasons can lead to substitutions, a larger share of product substitution in a location in the price data shows that the official authority believes that a larger share of the existing consumption basket needs to be replaced by new varieties to maintain its representativeness of household consumption, which likely suggests that more varieties are being introduced in those places. To account for the differences in consumption weights across product categories, I construct the following measure of municipality-level substitution, where j indexes municipalities:

$$\text{Weighted\_Substitution\_Share}_{jt,t-1} = \sum_i s_{i,t-1} \; \frac{N_{ijst,t-1}}{N_{ijst,t-1} + N_{ijut,t-1}} \tag{1.5}$$

where for municipality j, $N_{ijst,t-1}$ is the number of products substituted between period t and period $t-1$ in product category i, and $N_{ijut,t-1}$ is the number of products not substituted between period t and period $t-1$ in product category i. $s_{i,t-1}$ represents the expenditure shares on product category i.

Table 1.7 reports results of regressions with or without the full set of location-level controls. The dependent variable in columns 1 and 2 is the unweighted average number of substitutions within a municipality, which does not account for differences in expenditure share (i.e., importance) across product categories; in columns 3 and 4, I use the weighted substitution share variable defined earlier. Looking at column 4, which includes the full set of location-level controls, I find that a 1 percentage point increase in the growth rate of predicted local labor demand in a municipality leads to a 0.25 percentage point higher share of products that experienced substitution in the official catalog. This result provides another piece of evidence in support of the hypothesis that newer varieties are more frequently introduced in markets experiencing positive labor demand shocks, likely due to firms' response to the increase in consumers' purchasing power.

A decrease in markup among existing varieties due to competitive pressure from product innovation can also be propelled by changes in the mix of retail establishments. Large, modern supermarkets and department stores tend to feature a wider selection of products than traditional retail establishments; as such, an increase in modern retail establishments can reinforce the pro-competitive price effect. To evaluate this hypothesis, I examine the effect of the shift-share instrument on the number of establishments of different types of wholesale and retail businesses. Panel B of Table 1.6 reports the results. I find that an increase in predicted local labor demand has no effect on the number of wholesale and retail units overall, slightly decreases the number of food and beverage retail units (mostly mom-and-pop stores and market stalls), but increases the number of department stores, which may suggest that booming locations likely attracted a larger number of this kind of large modern retailers. This is another piece of evidence that points to decreased markups pushed down by an increase in varieties as the most likely mechanism that decreases tradables inflation in places experiencing higher income growth.

1.7 Model

Are the observed empirical results consistent with economic theory regarding consumer and firm behavior? The key empirical result is that a predicted local labor demand shock, which increases local market size by increasing local income, leads to lower inflation of continuing goods varieties in those locations, likely through increasing the rate of new product entry. I develop a general equilibrium model consistent with my findings by incorporating translog preferences and endogenous markups in a standard CES preference framework. The model allows me to conduct comparative statics exercises and demonstrate that an increase in market size leads to lower markups and lower prices on continuing varieties.

Model Intuitions. When local income increases due to a positive local labor demand shock, demand for products should increase. If the short-run supply curve for products is upward-sloping, the equilibrium price will go up. However, if the increase in demand leads to firms' endogenously responding to changes in market size by increasing product entry, the equilibrium price increase will be mitigated or even reversed. As more varieties become available, consumers spread out their expenditure across more products and spend less on each existing variety. Depending on the curvature of the utility function, lower spending on a variety can lead to higher elasticity of demand and therefore lower markups.

Prior literature has been receptive to the idea that demand is more elastic at lower quantities: For example, the seminal Krugman (1979) model assumes that the elasticity of demand is decreasing with respect to quantity consumed, i.e., $\frac{\partial \varepsilon}{\partial c} < 0$. Krugman argues that "this assumption, which might alternatively be stated as an assumption that the elasticity of demand rises when the price of a good is increased, seems plausible," and it is the key assumption behind the result whereby opening up trade leads to pro-competitive effects that lower markups. Mrázová and Neary (2017) refer to the condition $\frac{\partial \varepsilon}{\partial c} < 0$ as the "subconvexity" of demand, and point out that other authors, including Marshall (1920) and Dixit and Stiglitz (1979), have argued that subconvexity is intuitively more plausible than other forms of demand.[22] Few utility functions satisfy subconvexity and are also tractable at the same time: The translog utility function is one of them, which is why I include it in the model.

Appendix B provides a micro-foundation for the result whereby markup could fall as market size increases, using the general monopolistic competition model of Zhelobodko et al. (2012) with free entry. In this general model, in response to an increase in market size, the equilibrium markup may increase, decrease, or remain the same, depending on the curvature of the utility function.

Model Overview. The model closely follows work by Jaravel (2019) and Feenstra (2003). It features non-homothetic preferences of consumers across sectors and translog preferences within sectors on the demand side, and location-specific product entry decisions and monopolistic competition on the supply side. I solve for the equilibrium and show comparative statics in relation to market size. Since the model primarily seeks to guide interpretation of the negative effects on the inflation rate of tradable goods, the current version abstracts from wage-setting and the presence of a services sector. Future work will enrich the model by adding a wage-setting mechanism, a services sector, and the production of an "outside good" (i.e., exports) that is the source of increased local demand and income. The model is static to allow for tractable, closed-form solutions, but it can still yield insights for dynamic changes since we can consider changing from one time period to another in the data as moving from one equilibrium to another in the model.

Consumers

Nonhomothetic CES across sectors. There are I locations in this economy indexed by i. Each location features L_i number of households that consume and produce in K different sectors of the economy. All consumers within location i are homogeneous. The number of varieties available in each location and sector is endogenous and denoted N_{ik}. Households maximize aggregate consumption C_i, which combines consumption in each sector with a nonhomothetic constant elasticity of substitution (CES) aggregator:

$$C_i = \left(\sum_{k=1}^{K} \Omega_{ik}^{\frac{1}{\sigma_i}} \left(C_{ik} \right)^{\frac{\sigma_i - 1}{\sigma_i}} \right)^{\frac{\sigma_i}{\sigma_i - 1}} \tag{1.6}$$

[22] They also point out that subconvexity is consistent with much of the available empirical evidence on proportional pass-through, which suggests that it is less than 100% (Gopinath and Itskhoki, 2010; De Loecker et al., 2016).

where location-sector specific weights Ω_{ik} capture nonhomotheticities. σ_i is the location-specific elasticity of substitution across sectors. Standard CES results imply that consumption in each sector k is given by

$$C_{ik} = \sum_k \left(\frac{P_{ik}}{P_i}\right) C_i$$

where P_{ik} is the sectoral price index in location i and P_i is the aggregate CES price index in location i:

$$P_i = \frac{E_i}{C_i} = \left(\sum_{k=1}^{K} \Omega_{ik} P_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{1.7}$$

where $E_i = \sum_{k=1}^{K} P_{ik} C_{ik}$ is the expenditure of a consumer in location i. I can also define expenditure shares of consumers in location i on goods in sector k as

$$s_{ik} = \frac{P_{ik} C_{ik}}{E_i} = \frac{\Omega_{ik} P_{ik}^{1-\sigma}}{\sum_{k=1}^{K} \Omega_{ik} P_{ik}^{1-\sigma}} \tag{1.8}$$

Translog preferences within sectors. Each sectoral consumption k is itself a consumption aggregator. Within a sector, all consumers have the same translog preferences. Translog preferences do not have an explicit utility function representation, but it is straightforward to write out the expenditure function in terms of prices and elasticities. Let $\tilde{N}_{ik}$ be the total number of varieties conceivably available; this is the upper bound of the endogenous number of varieties introduced to location i and sector k, and I treat this number as fixed. Within each location i and sector k, the expenditure function is defined as

$$\ln(E_{ik}) = \ln(U) + \alpha_0 + \sum_{n=1}^{\tilde{N}_{ik}} \alpha_n \ln(p_{ink}) + \frac{1}{2} \sum_{n=1}^{\tilde{N}_{ik}} \sum_{m=1}^{\tilde{N}_{ik}} \gamma_{nm} \ln(p_{ink}) \ln(p_{imk}) \tag{1.9}$$

with $\gamma_{mn} = \gamma_{nm}$ for all m, n. I impose the restriction $\sum_{n=1}^{\tilde{N}_{ik}} \alpha_n = 1$ and $\sum_{n=1}^{\tilde{N}_{ik}} \gamma_{nm} = 0$ to ensure that the expenditure function is homogeneous of degree one. Following the convention in the literature, I simplify the above function by assuming that all varieties enter symmetrically into the expenditure function:

$$\alpha_n = \frac{1}{\tilde{N}_{ik}}, \gamma_{nn} = \frac{-\gamma_{ik}(\tilde{N}_{ik} - 1)}{\tilde{N}_{ik}}, \text{ and } \gamma_{mn} = \frac{\gamma_{ik}}{\tilde{N}_{ik}} \text{ for } m \neq n, m, n = 1, \cdots, \tilde{N}_{ik}$$

Feenstra (2003) shows that in the symmetric case, the expenditure function can be further simplified if I assume $\gamma > 0$ and that only the goods $n = 1, \cdots, N$ are available, so that I do not need to worry about the price of unavailable goods. Dropping subscripts i and k for convenience, equation 1.9 simplifies to

$$\ln(E) = \ln(U) + a_0 + \sum_{n=1}^{N} \alpha_n \ln(p_n) + \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} b_{nm} \ln(p_n) \ln(p_m) \tag{1.10}$$

where $\alpha_n = \frac{1}{N}, b_{nn} = \frac{-\gamma(N-1)}{N}$, and $b_{mn} = \frac{\gamma}{N}$ for $m \neq n$, with $m, n = 1, \cdots, N$, and $a_0 = \alpha_0 + \frac{1}{2}\frac{\tilde{N}-N}{\gamma\tilde{N}N}$. By Shephard's Lemma, the spending share on each variety n is

$$s_n = \frac{1}{N} + \gamma\left(\left[\frac{1}{N}\sum_{m=1}^{N}\ln(p_m)\right] - \ln(p_n)\right) \tag{1.11}$$

I can also derive the elasticity of demand for each variety n as

$$\epsilon_n = 1 - \frac{d\ln(s_n)}{d\ln(p_n)} = 1 + \frac{(N-1)\gamma}{Ns_n} \tag{1.12}$$

Expression 1.12 shows that the elasticity of demand depends on both consumer's expenditure share and the number of varieties. Elasticity for a particular variety is higher when the consumer spends less on it or when there are many close substitutes available.

Labor supply and budget constraint. I assume each household supplies l_i units of labor inelastically. The wage for one unit of labor is the numeraire.[23] Each household in location i is subject to the following budget constraint:

$$\sum_{k=1}^{K}\sum_{n=1}^{N_{ik}} c_{ink}p_{ink} = l_i \tag{1.13}$$

Producers

Monopolistic competition between homogeneous firms. I assume that within each sector k and location i, symmetric firms compete monopolistically.[24] Firms maximize the sum of profits $\sum_i \pi_{ik}$ across locations i. In each location i, the quantity produced by the firm is $q_{ik} = Q_{ik}l_{ik}$, where l_{ik} is local labor demanded for production and Q_{ik} is a productivity factor. There is also a τ_{ik} per unit cost in monetary terms, which captures trade costs and costs of manufacturing at the source location. To enter location i, all firms must also pay a fixed cost of f'_{ik} effective units of labor, so the required amount of labor for entry into location i is $\frac{f'_{ik}}{Q_{ik}}$. This fixed cost reflects various costs for entering a new location: setting up retail locations or establishing a partnership with retailers, advertising, etc. The variable cost reflects the cost of selling the variety locally: hiring local labor, arranging transport, etc.[25]

In this model, local labor is not used to manufacture the varieties since our focus is on the local retail response to changes in market size. The majority of tradable goods

[23] The empirical sections show that an increase in predicted local labor demand also increases local wages. However, the model primarily serves to illustrate the endogenous markup response and thus abstracts from wage responses. Future iterations of the model will allow for location-specific wages.

[24] In the model, there is a one-to-one relationship between a firm and a variety. Jaravel (2019) shows that a model of similar setup is robust to the introduction of multi-product firms.

[25] In my model, the firm sells directly to consumers. I am working on an extension that includes both retailers and wholesalers.

in the consumption basket, such as produce, clothing, cosmetics, electronics, vehicles, and entertainment equipment, are not produced locally: They are either imported or produced in other municipalities in Mexico. Furthermore, even if some items are locally produced, the empirical results (in which I exclude the sector that produces the goods in constructing of the shift-share instrument) show that the observed lower inflation in places experiencing a positive labor demand shock is unlikely to be driven by location-specific productivity shocks (i.e., changes in Q_{ik}). Thus, I believe it is reasonable to exclude local labor in the model for better tractability, but future work will include the production of goods within the location as a part of the model.

The marginal cost of selling each unit is $1/Q_{ik} + \tau_{ik}$. For simplicity, let

$$Z_{ik} = \frac{Q_{ik}}{1 + \tau_{ik} Q_{ik}}, \text{ and } f_{ik} = \frac{f'_{ik}}{1 + \tau_{ik} Q_{ik}}$$

Then the marginal cost equals $1/Z_{ik}$. The profit of each firm in sector k obtained by selling in location i is given by

$$\pi_{ik}(q_{ik}) = p_{ik}(q_{ik})q_{ik} - \frac{q_{ik}}{Z_{ik}} - \frac{f_{ik}}{Z_{ik}}$$

Because the firm's profit maximization problem is additively separable, it is equivalent to each firm's branch in each location i maximizing location-specific profits $\pi_{ik}(q_{ik})$.[26] The total amount of labor required by all firms in sector k and location i is

$$L_{ik} = N_{ik}\left(\frac{q_{ik}}{Z_{ik}} + \frac{f_{ik}}{Z_{ik}}\right)$$

Markups. Firms engage in monopolistic competition, so they set their markup equal to $1/(\epsilon_{ik} - 1)$. By equation (1.12) and the fact that $s_{ink} = \frac{1}{N_{ink}}$ by symmetry, I have

$$M^*_{ik} = \frac{1}{\epsilon_{ik} - 1} = \frac{1}{(N_{ik} - 1)\gamma_{ik}} \tag{1.14}$$

We can see that as the number of varieties (or alternatively, firms) increases in a location, the markup decreases and will eventually drop to zero.

Free entry. Each firm's marginal cost of selling in location i is $1/Z_{ik}$. As a result, the price each firm charges given the optimal markup is

$$p^*_{ik} = \frac{1 + M^*_{ik}}{Z_{ik}} \tag{1.15}$$

Firms enter location i and sector k until profits are brought to zero via decreasing markups. Using the fact that the wage is the numeraire, the free-entry condition can be written as

$$\pi_{ik} = M^*_{ik}\frac{q_{ik}}{Z_{ik}} - \frac{f_{ik}}{Z_{ik}} = 0 \tag{1.16}$$

[26] One way to think about this is that retail operations are delegated to franchises across locations. The franchises make decisions independently, which also maximizes the total profit of the firm, given that there are no spillovers across locations.

Equilibrium

Equilibrium conditions are as follows:

1. Optimal allocation of spending across sectors: equation (1.8)
2. Budget constraint: equation (1.13)
3. Optimal markups: equation (1.14)
4. Zero-profit condition: equation (1.16)
5. Product market clearing: The supply of each variety n in each sector k and location i equals demand $q_{ik} = L_i c_{ik}$ (omitted subscript n by symmetry)
6. Labor market clearing: Labor demand equals labor supply in each location i

$$\sum_k \frac{N_{ik}}{Z_{ik}}(L_i c_{ik} + f_{ik}) = L_i l_i$$

Proposition 1. The equilibrium across locations i and sectors k is characterized by

$$\text{Number of varieties } \forall i,k : N_{ik}^* = \frac{(\gamma_k - 1 + \sqrt{(1-\gamma_{ik})^2 + 4\gamma_{ik}\frac{L_i l_i s_{ik} Z_{ik}}{f_{ik}}})}{2\gamma_{ik}} \tag{1.17}$$

$$\text{Sectoral Price Index } \forall i,k : \ln(P_{ik}^*) = \alpha_{0k} + \frac{1}{2}\frac{1-\frac{N_{ik}^*}{\tilde{N}_{ik}}}{\gamma_{ik}N_{ik}^*} + \frac{1+(N_{ik}^*-1)\gamma_{ik}}{(N_{ik}^*-1)\gamma_{ik}}\frac{1}{Z_{ik}} \tag{1.18}$$

$$\text{Aggregate Price Index } \forall i : P_i^* = \left(\sum_{k=1}^{K} \Omega_{ik} P_k^{1-\sigma}\right)^{\frac{1}{1-\sigma}} \tag{1.19}$$

Proof. I first solve for the equilibrium number of varieties N_{ik}^*. From the free-entry condition (1.16) and the product market clearing condition,

$$M_{ik}^*(L_i c_{ik}) = f_{ik}$$

Equation (1.8) and within-sector symmetry allow us to derive the equilibrium spending on each variety within each sector k and location i to be

$$e_{ik} = p_{ik}(L_i c_{ik}) = L_i l_i \frac{s_{ik}}{N_{ik}}$$

where $L_i l_i$ is total income in location i and s_{ik} is the share of income in location i spent on goods from sector k.

Using the free-entry condition (1.16) and the two equations derived above, I can write

$$\begin{aligned} M_{ik}^* \frac{e_{ik}}{p_{ik}} &= f_{ik} \\ \frac{1}{(N_{ik}-1)\gamma_{ik}} \frac{L_i l_i s_{ik}}{\frac{N_k}{Z_{ik}} \frac{1+(N_{ik}-1)\gamma_{ik}}{(N_{ik}-1)\gamma_{ik}}} &= f_{ik} \\ \frac{L_i l_i s_{ik} Z_{ik}}{N_{ik}(1+(N_{ik}-1)\gamma_{ik})} &= f_{ik} \\ N_{ik}^2 \gamma_{ik} + (1-\gamma_{ik})N_{ik} - \frac{L_i l_i s_{ik} Z_{ik}}{f_{ik}} &= 0 \end{aligned}$$

Thus, the equilibrium number of varieties is

$$N_{ik}^* = \frac{(\gamma_k - 1 + \sqrt{(1-\gamma_{ik})^2 + 4\gamma_{ik}\frac{L_i l_i s_{ik} Z_{ik}}{f_{ik}}})}{2\gamma_{ik}} \tag{1.20}$$

I also write out the equilibrium price of each variety p_{ik}^*:

$$p_{ik}^* = \frac{1+(N_{ik}-1)\gamma_{ik}}{(N_{ik}-1)\gamma_{ik}} \frac{1}{Z_{ik}} \tag{1.21}$$

Following Feenstra (2003) and using the symmetry of all varieties, the aggregate price index within each sector is given by

$$\ln(P_{ik}^*) = \alpha_{0k} + \frac{1}{2} \frac{1 - \frac{N_{ik}^*}{\tilde{N}_{ik}}}{\gamma_{ik} N_{ik}^*} + \frac{1+(N_{ik}^*-1)\gamma_{ik}}{(N_{ik}^*-1)\gamma_{ik}} \frac{1}{Z_{ik}} \tag{1.22}$$

Equation (18) shows that as the number of varieties N_{ik}^* increases, the sector-level price index goes down for two reasons. First, consumers exhibit a love of variety, captured by the term $\frac{1}{2}\frac{1-\frac{N_{ik}^*}{\tilde{N}_{ik}}}{\gamma_{ik} N_{ik}^*}$. Second, an increasing number of varieties exerts competitive pressure on existing varieties and leads to lower markups, which is captured by the term $\frac{1+(N_{ik}^*-1)\gamma_{ik}}{(N_{ik}^*-1)\gamma_{ik}} \frac{1}{Z_{ik}}$ (note that $\frac{1+(N_{ik}^*-1)\gamma_{ik}}{(N_{ik}^*-1)\gamma_{ik}} = 1 + M_{ik}^*$). The empirical results primarily speak to effects that operate through the second channel of decreasing relative prices on existing varieties. A future exercise would be to quantify the full price index implications of local labor demand shocks in Mexico by constructing price indices that account for product entry and exit (e.g., the Sato-Vartia exact CES price index).

Comparative statics

The empirical result shows that an increase in predicted local labor demand leads to lower inflation of continuing varieties among tradable items. This can be shown directly in the model by taking comparative statics with respect to the number of consumers L_i and income per capita l_i. The predicted local labor demand shock increases both employment and wages per worker, so both L_i and l_i are increasing.

Equation (1.20) shows that changes in the number of consumers (L_i) and in income per capita (l_i) have the same effect on the equilibrium. From equations (1.20), (1.21), (1.21) and (1.7), the comparative statics of interest are as follows.

$$\text{Variety increases with demand: } \frac{dN_{ik}^*}{dL_i} = \frac{dN_{ik}^*}{dl_i} > 0 \ \forall k, i$$
$$\text{Markup decreases with demand: } \frac{dM_{ik}^*}{dL_i} = \frac{dM_{ik}^*}{dl_i} < 0 \ \forall k, i$$
$$\text{Prices decrease with demand: } \frac{dp_{ik}^*}{dL_i} = \frac{dp_{ik}^*}{dl_i} < 0 \ \forall k, i$$
$$\text{Sectoral price index decreases with demand: } \frac{dP_{ik}^*}{dL_i} = \frac{dP_{ik}^*}{dl_i} < 0 \ \forall k, i$$
$$\text{Aggregate price index decreases with demand: } \frac{dP_{i}^*}{dL_i} = \frac{dP_{i}^*}{dl_i} < 0 \ \forall i$$

These comparative statics show that when either the number of consumers or income per capita increases in a location, the total number of varieties available in that location increases, the price of each variety decreases, the sectoral price index decreases, and the aggregate price index decreases. This yields a downward-sloping supply curve, consistent with the empirical results.

1.8 Price index quantification

Finally, I study the price index and inequality implications of the findings by quantifying, through a back-of-the-envelope calculation, the effects of a local labor demand shock on changes in the local price index faced by households in different locations and different quantiles of the income distribution. Changes in the price index faced by a representative household in location j and income quantile q may differ from those faced by another household in location j' and quantile q' because (1) the changes in local labor demand in location j may differ from j', (2) the expenditure shares on product categories for households in location j may differ from location j', and (3) even in the same location, the expenditure shares on product categories for households in quantile q may differ from those of households in quantile q' (for example, poorer households spend a higher share of their income on food and less on services). The main results show that a 1 percentage point increase in the growth rate of local labor demand leads to a 0.84 percentage point increase in the services inflation rate, but a 0.07 percentage point decrease in the goods inflation rate. In light of the heterogeneity of the price effect for different spending categories, I perform an exercise that asks: How much does local inflation change if the growth rate of local labor demand increases by 10 percentage points, which corresponds to moving from the median municipality to the 90th percentile municipality in the sample. The variable of interest is local non-housing inflation, measured by a simple Laspeyres price index, where j indexes metropolitan areas and t indexes years:

$$P_j^{t,t-5} = \sum_{i \in I} \frac{p_{i,t}}{p_{i,t-5}} s_{i,t-5} \tag{1.23}$$

In contrast to Equation (1.4), here i denotes product groups (e.g., processed food, unprocessed food, clothing, energy, etc.) for which I have estimates based on the heterogeneity analysis conducted above. I exclude housing since I do not have reliable estimates of the effect of labor demand shocks on housing prices in Mexico, given that housing prices are measured with substantial noise in the CPI data. I do observe a positive albeit statistically insignificant estimate on housing prices, which is consistent with our common intuition that labor demand shocks increase the cost of housing. As a result, the effect on the overall price index should be higher than that on the non-housing prices.

I use household expenditure data from the 2000 ENIGH survey to construct shares $s_{i,t-5}$. I chose the 2000 ENIGH survey because it supplies product-category weights for the CPI data from 2002 to 2010, which covers the midpoint of my study period (Gagnon, 2018). I construct metropolitan area-specific household income quintiles based on households' monthly income reported on the ENIGH survey, and calculate metropolitan area and income quintile-specific expenditure shares across 22 product groups by matching the spending of each household on a specific item to the product group it belongs to in the CPI data.[27] Since the CPI data from 2002 to 2010 use expenditure shares from the 2000 ENIGH survey as weights, the matching between items in ENIGH and product groups in the CPI data is straightforward. For the price changes $\frac{p_{i,t}}{p_{i,t-5}}$, I plug in product group-specific regression coefficients from my empirical analysis.

Table 1.8 presents summary statistics of the quantification exercise. Panel A shows the average household expenditure share on goods and services as well as the effect of a 10 percentage point increase in the growth rate of local labor demand on local non-housing inflation, across the 46 metropolitan areas in the sample. On average, households spend 46% of their income on goods and 22% on services. A 10 percentage point increase in the growth rate of local labor leads to a 0.017 percentage point lower non-housing inflation. Although there is some geographic variation, the difference is not substantial. The small effect suggests that the negative effect on goods inflation and the positive effect on services inflation mostly cancel out each other.

Panel B performs the same exercise as in Panel A, but splits the sample across five income quintiles. Quintile 1 consists of the 20% of households with the lowest income in each metropolitan area and quintile 5 consists of the richest 20%. Within each metropolitan

[27] I use metropolitan areas instead of municipalities in this exercise to increase statistical power, since there may be zero or very few households surveyed in certain municipalities. At metropolitan area level, the average number of households is 105, the median is 85.5, the 10th percentile is 22, and the 90th percentile is 165.

area, rich households spend a higher share of their income on services and a lower share on goods than poorer households: The richest 20% of households in each metropolitan area spend 25% of their expenditure on services and 41% on goods, whereas the poorest 20% spend 15% on services and 51% on goods. As a result, local labor demand shocks increase inflation for rich households but decrease inflation for poor households: Given a 10 percentage point increase in the growth rate of local labor demand, the richest 20% of households in each metropolitan area would experience a 0.084 percentage point increase in the non-housing inflation rate, but the poorest 20% of households would experience a 0.154 percentage point *decrease* in the non-housing inflation rate. Although the estimates are not large in magnitude, they nonetheless point to the potential for positive local labor market shocks in reducing within-location inequality between richer and poorer households by slowing the inflation of products that poorer households consume more of.

Figure 1.6 visualizes changes in the price index in response to a 10 percentage point increase in the growth rate of local labor demand over a five-year period for households in the 1st local income quintile (poorest) and those in the 5th local income quintile (richest). The blue color denotes places with the lowest inflation and the red color denotes those with the highest inflation. Subfigure (a) shows that low-income households cities in northern Mexico close to the US border would see their price index increase (or decrease by a smaller amount than those in the center and south of Mexico). This is because the poorest households in northern Mexico—a region that has experienced dramatic increases in economic growth since the 1990s—are on average richer than poor households in other metropolitan areas, and tend to spend a higher share of their income on services. A local labor demand shock increases services inflation but decreases goods inflation, so households that spend more on services will see their price index increase more. Subfigure (b) shows the results for the richest 20% of households in each metropolitan area: here the patterns are reversed. Rich households in Mexico City, the center of the country, and a few pockets in the east and south (Veracruz, Villahermosa) would experience the highest inflation from a labor demand shock, while rich households in the north would see less inflation. Even though the north of Mexico is richer than other regions, high-income households there are not as rich and do not spend quite as much on services as high-income households in some other cities.

Taken together, the results suggest that within cities, local labor demand shocks can slightly reduce inequality because it decreases the inflation of goods that are disproportionately consumed by poorer households. On average, geographic variations in consumption patterns are not high across metropolitan areas in Mexico, but within income quintiles there is significant variation in spending patterns that cause differential effects from a local labor demand shock of a given magnitude.

1.9 Conclusion

The question of how shocks to local labor markets affect real income has been a prominent academic and policy topic in many fields of economics. To help understand the implications of shocks for real income—nominal income divided by the price index—it is crucial to account for the local price effects of shocks because price changes may offset or reinforce nominal income gains. Two central questions in this context are (a) how a positive local labor demand shock affects the inflation of different kinds of goods and services and (b) whether the observed effects on the price index have reduced or exacerbated inequality, both across geography and the income distribution.

This paper provides empirical evidence that informs both questions. To do so, I combine a rich dataset of barcode-by-store-level price quotes for consistently sampled products with firm-level data on wages and employment across 15 years in Mexico. I find that a 1 percentage point increase in the growth rate of municipality-level local labor demand leads to a 0.061 percentage point increase in the local services inflation rate, but a 0.099 percentage point decrease in the local goods inflation rate. The negative results on tradable goods are consistent with a mechanism whereby an increase in market size drives new product entry, and the increased competitive pressure on existing products decreases their markups. Comparative statics from a general equilibrium model with endogenous markups and higher demand elasticity at lower consumption levels provide theoretical support for the empirical results. Conversely, the positive results on services are likely due to services prices' responding more to changes in local wage and consumer demand being less elastic for services. For a labor demand shock of a given magnitude, low-income households would experience lower inflation than high-income households because the latter spend less on goods and more on services.

The specific context of Mexico—a middle income country experiencing trade liberalization but with still not well-integrated local markets—likely plays an important role in driving the results. Many multinational retail chains began their operations in Mexico in the 1990s and only started significantly expanding in the 2000s. Internal trade costs are high due to poor infrastructure and spatial inequality is prominent, both of which create conditions for pricing at the local level according to local demand conditions. In the US, DellaVigna and Gentzkow (2019) find that retail chains tend to price uniformly for the same item sold at different locations, due to primarily managerial inertia. In a setting where mom-and-pop stores, market vendors, franchises, and authorized resellers constitute a significant portion of the retail landscape, firms are more fragmented across space and therefore are more likely to price according to location-specific demand conditions. Studies in low-income setting have also found quantitatively small overall price index effects from income shocks (cash transfers), but the mechanisms there are different: Egger et al. (2022) find that there is significant slack labor capacity in rural Western Kenya, and as a result, increasing aggregate demand produces small price effects because labor is now being utilized more intensively for the same amount of wages paid. Taken together, the results from this paper point to the important role of local market and economic conditions in determining the evolution of local prices, and have broad implications for our understanding of place-based policies, variable markups, and the consequences of globalization.

1.10 Figures and Tables

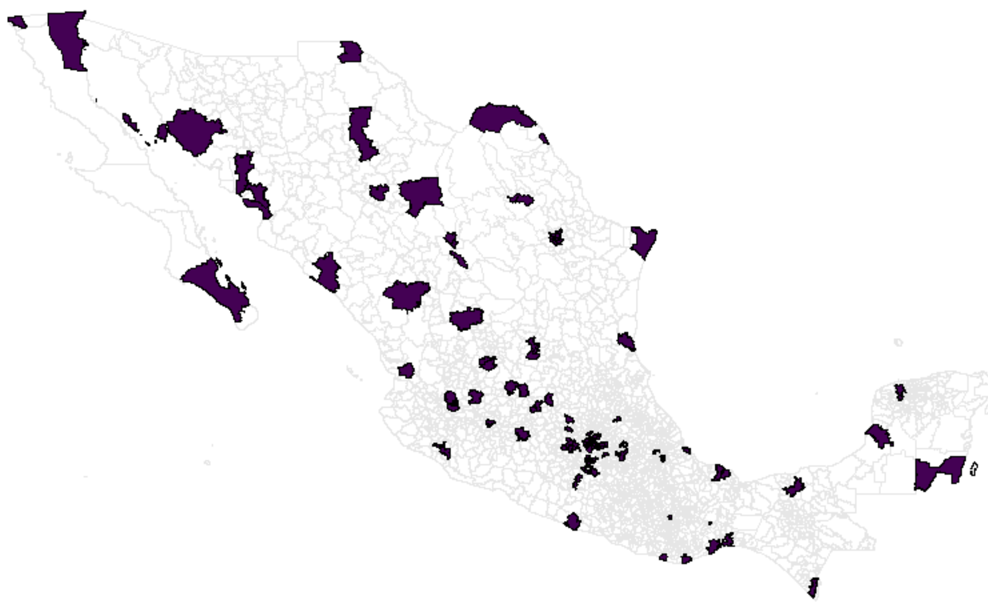
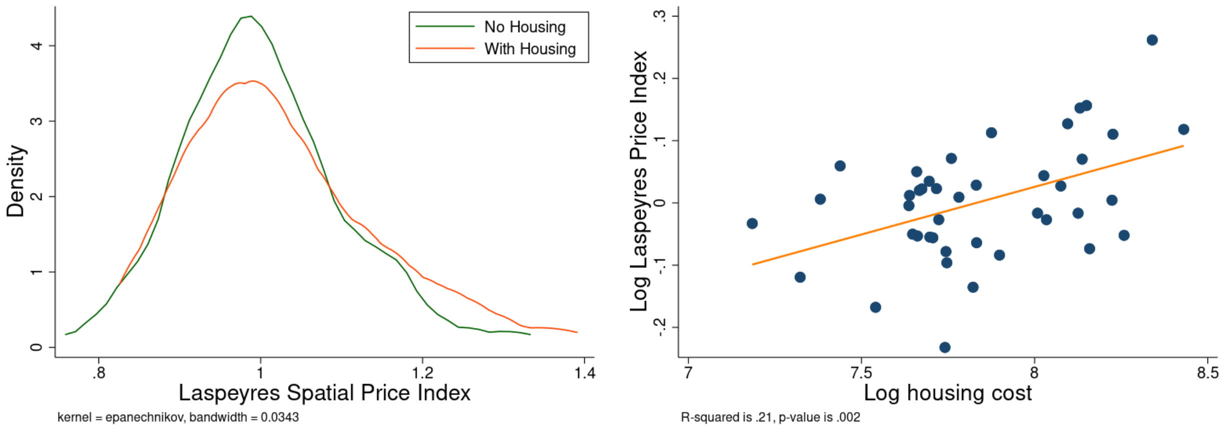


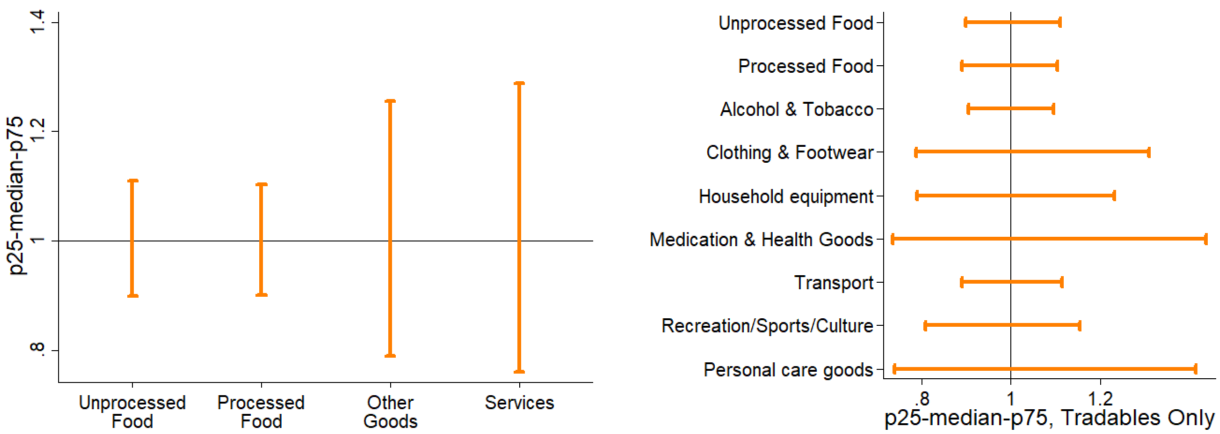
Figure 1.1: Map of Mexican municipalities where price quotes were collected



(a) Distribution of Spatial Price Indices (b) Non-housing Price Index and Housing Cost

Figure 1.2: Spatial price indices

Note: Figure (a) plots the kernel density of the Laspeyres spatial price index across metropolitan areas, for both the index including housing costs (calculated from ENIGH data) and excluding housing costs. Figure (b) shows the scatter plot between log housing cost and the Laspeyres spatial price index.

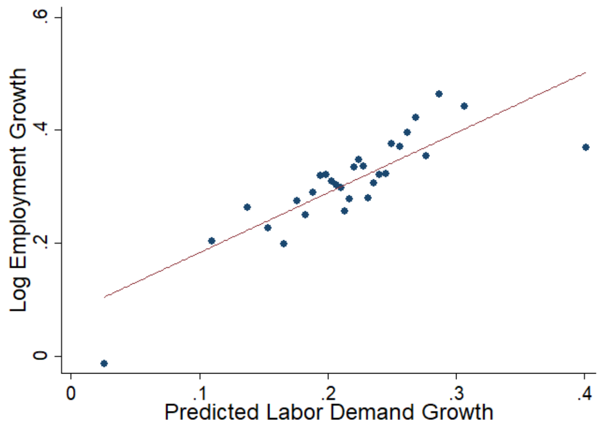


(a) Across four broad groups

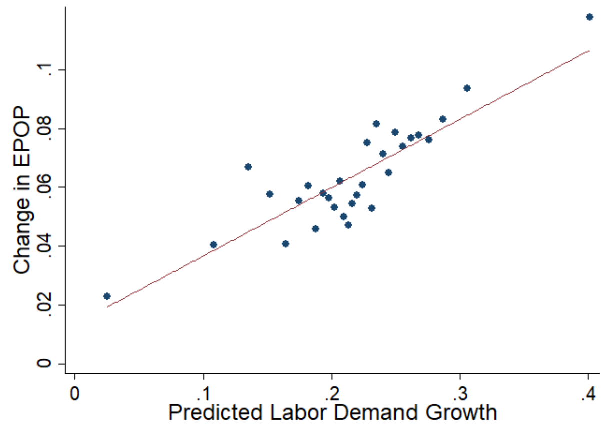
(b) Tradables only, across finer COICOP groups

Figure 1.3: Spatial dispersion of prices

Note: Figure (a) shows the interquartile range (p25-p75) of the average prices across metropolitan areas of four broad categories of goods and services that partition the CPI data relative to the median, which is standardized to 1. Figure (b) shows the interquartile range for 8 product groups within tradables that follow the Classification of Individual Consumption According to Purpose (COICOP).



(a) Log employment growth



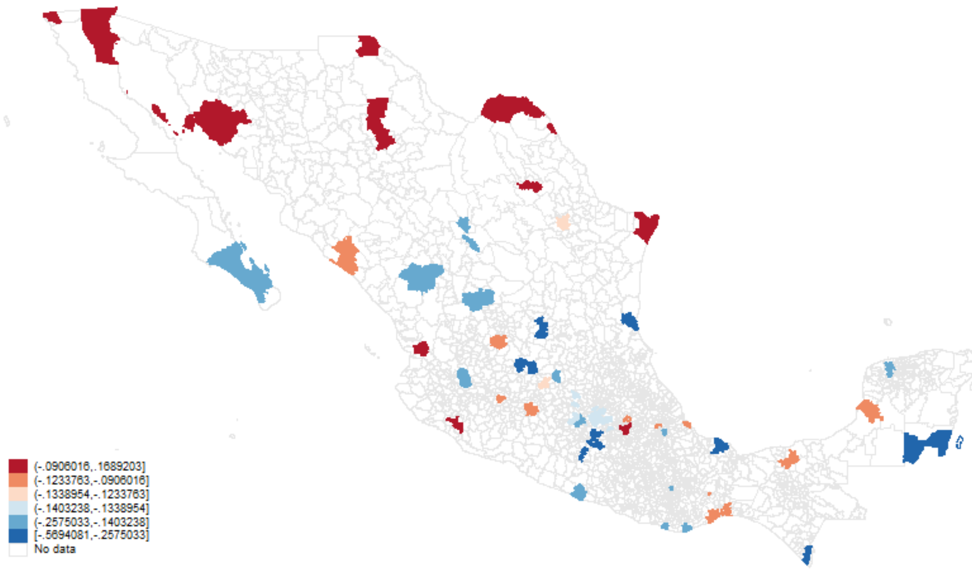
(b) Employment-to-population ratio

Figure 1.4: Correlation between the shift-share IV and actual employment growth

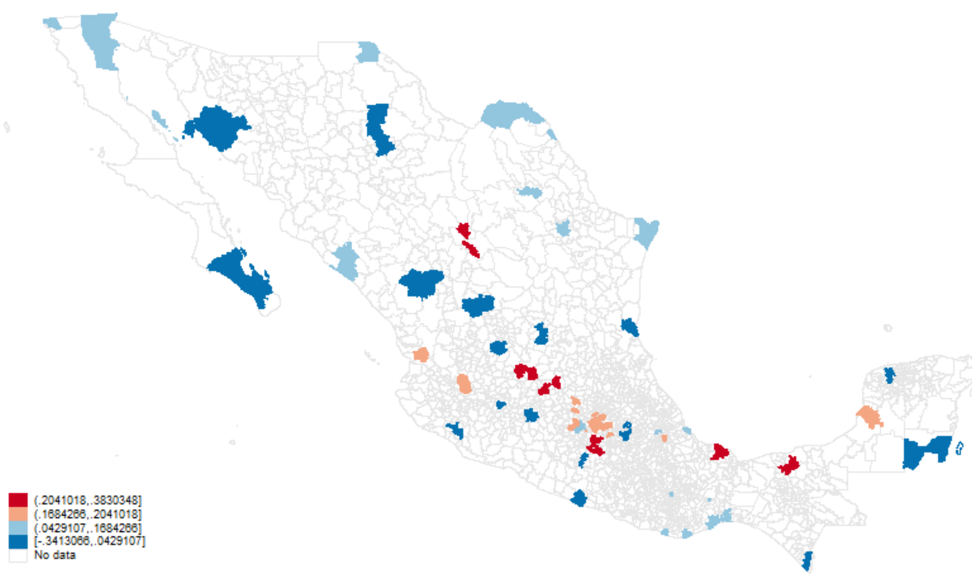
Note: The figure shows binned scatter plots between (a) log employment growth and (b) changes in the employment-to-population ratio and the shift-share instrument that measures labor demand growth at municipality-period level, after absorbing year fixed effects. The sample includes all municipalities in Mexico. N=7,257.

Figure 1.5: Heterogeneity across tradable product categories

Note: Coefficients from a regression of log price changes on the shift-share instrument, where the instrument is interacted with product group dummies. The regression controls for lagged location-level variables specified in Table 1.2. Product groupings are based on the classification from Gagnon (2009) according to the Classification of Individual Consumption According to Purpose (COICOP).



(a) Poorest Quintile



(b) Richest Quintile

Figure 1.6: Effects of a 10 pp increase in labor demand on local non-housing inflation

Note: The figure shows the effect on the local non-housing inflation rate for (a) the poorest 20% of households in each metropolitan area and (b) the poorest 20% of households in each metropolitan area. See Table 1.8 notes for details on the methods.

Table 1.1: Shift-share shock balance tests

Panel A: Pre-trends Test

Balance Variable	Coef	SE
Price growth, 1993-1998 (all items)	-0.178	(0.154)
Price growth, 1993-1998 (goods)	-0.077	(0.181)
Price growth, 1993-1998 (services)	-0.317	(0.181)
Price growth, 2003-2008 (all items)	-0.267	(0.155)
Price growth, 2003-2008 (goods)	-0.104	(0.092)
Price growth, 2003-2008 (services)	-0.866	(0.468)

Panel B: Industry-level Balance

Balance Variable	Coef	SE
Lagged log labor income	0.060	(0.053)
Lagged log employment	-0.068	(0.049)
Lagged log value added	0.180	(0.121)
Sector's average employment share	-0.002	(0.002)
Number of industry-periods	241	

Panel C: Location-level Balance

Balance Variable	Coef	SE
Completed secondary school	0.075	(0.061)
Completed university	0.035	(0.037)
Share of women in workforce	0.107	(0.021)
Share of adults speaking an indigenous language	-0.060	(0.013)
Share of employment in manufacturing	0.220	(0.048)
Share of employment in services	0.027	(0.057)
Number of location-periods	270	

Note: Panel A reports results from regressing five-year price changes on the next five-year changes in predicted local labor demand to test for pre-trends. Panel B reports coefficients from industry-level regressions of start-of-period industry characteristics from the Economic Census on industry-level shocks, controlling for period indicators and weighting by average industry exposure shares. Panel C reports coefficients from regressions of location-level (metropolitan area for 1993-1998, municipality for 2003-2008 and 2013-2018) characteristics on the shift-share instrument, controlling for period indicators. Exposure-robust standard errors are reported in parentheses following Borusyak, Hull, and Jaravel (2022).

Table 1.2: Effect of local shocks on local price changes

Reduced-form regression:								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Predicted local labor demand)	-0.00538	0.166***	-0.0574*	0.0702**	-0.0991***	0.0611*	-0.0865***	0.0876*
	(0.0310)	(0.0518)	(0.0322)	(0.0304)	(0.0298)	(0.0361)	(0.0293)	(0.0480)
IV regression:								
Δ (Log wage bill per worker)	-0.0182	0.761*	-0.183**	0.260	-0.183***	0.118	-0.171***	0.169
	(0.102)	(0.387)	(0.0774)	(0.160)	(0.0462)	(0.0807)	(0.0516)	(0.112)
Observations	85,867	3,445	64,373	18,049	64,373	18,049	64,373	18,049
IV F-stat	14.76	5.632	15.45	12.46	45.03	29.09	41.75	27.13
Controls					X	X	X	X
Leave-one-out							X	X

Note: Each observation is a barcode-by-store-level product. Price changes are differences in the 3-month averages across three periods: 1993-1998, 2003-2008, 2013-2018. Standard errors are clustered at metropolitan area level. Bartik shift-share instrumental variables are constructed using 3-digit industry codes (89 unique values) at municipality level for 2003-08 and 2013-18 and metropolitan-area level for 1993-98. All regressions control for year×product category fixed effects. Housing is excluded. Columns 5-8 control for lagged variables of log wages, share completing secondary school, share completing university, share of women in workforce, share speaking an indigenous language, share of employment in agriculture, manufacturing, and services. In Columns 7 and 8, the shift-share instrument for each product category leaves out the 3-digit industry that produces those goods or services.

Table 1.3: Effect of local shocks on the local Laspeyres price index

Reduced-form regression:						
	(1)	(2)	(3)	(4)	(5)	(6)
Products:	All	All	Goods	Goods	Services	Services
Δ (Predicted local labor demand)	-0.0108	-0.00957	-0.0474**	-0.0479**	0.0510	0.0547
	(0.0303)	(0.0301)	(0.0219)	(0.0219)	(0.0432)	(0.0384)
IV regression:						
Δ (Log wage bill per worker)	-0.0214	-0.0195	-0.117***	-0.122***	0.102	0.116
	(0.0570)	(0.0581)	(0.0350)	(0.0346)	(0.101)	(0.0971)
Number of industry-periods	241	241	241	241	241	241
IV F-stat	19.05	17.61	16.17	14.63	14.64	14.21
Controls		X		X		X

Note: See Table 1.2 for detailed notes on variable construction and sample inclusion. Dependent variable is the municipality-level Laspeyres price index. Housing is excluded. Columns 2, 4, and 6 control for variables used in Column 5-8 of Table 1.2. Exposure-robust standard errors are reported in parentheses and obtained from equivalent industry-level IV regressions following Borusyak, Hull, and Jaravel (2022).

Table 1.4: Controlling for local public expenditure

Reduced-form regression:			
	(1)	(2)	(3)
Specification:	Baseline	Investment	Total
Δ (Predicted local labor demand)	-0.0991***	-0.0982***	-0.0987***
	(0.0298)	(0.0294)	(0.0304)
IV regression:			
Δ (Log wage bill per worker)	-0.183***	-0.181***	-0.182***
	(0.0462)	(0.0464)	(0.0468)
Observations	64,373	61,975	63,679
IV F stat	45.03	47.12	45.95
Controls	X	X	X

Note: See Table 1.2 for detailed notes on variable construction and sample inclusion. All regressions control for year×product category fixed effects and control for lagged variables of log wages, share completing secondary school, share completing university, share of women in workforce, share speaking an indigenous language, share of employment in agriculture, manufacturing, and services. Column 2 additionally controls for changes in municipality (metropolitan area for 1993-1998)-level log changes in public investment expenditure. Column 3 additionally controls for changes in municipality (metropolitan area for 1993-1998)-level log changes in total public expenditure.

Table 1.5: Effect of local shocks on price changes, substituted items

Reduced-form regression:

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Predicted labor demand)	0.150	0.0808	0.223	0.0424	0.278*	0.0611	0.285	0.176
	(0.110)	(0.165)	(0.154)	(0.197)	(0.161)	(0.176)	(0.171)	(0.202)
IV regression:								
Δ (Log wage) bill per worker)	0.529	0.434	0.786	0.140	0.710	0.198	0.822	0.584
	(0.397)	(0.902)	(0.523)	(0.657)	(0.477)	(0.590)	(0.553)	(0.790)
Observations	75,316	315	66,314	8,687	66,313	8,687	66,313	8,687
IV F-stat	19.94	3.274	19.72	15.34	24.51	8.500	22.52	8.053
Controls					X	X	X	X
Leave-One-Out							X	X

Note: See Table 1.2 for detailed notes, all of which apply here as well. The sample consists of items (barcode-by-store-level products) that experienced a change in item characteristics during a five-year period.

Table 1.6: Effect of local labor demand shocks on retail employment and establishments

Panel A: Percentage change in employment

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	Wholesale and Retail	All Retail	Food and Beverage	Convenience and Mini Stores	Department Stores	All Other Retail
Δ (Predicted local labor demand)	-0.0657	-0.121	-0.260**	0.531	1.125**	-0.0886
	(0.147)	(0.169)	(0.124)	(0.405)	(0.501)	(0.200)
Panel B: Percentage change in the number of establishments						
Δ (Predicted local labor demand)	0.0188	0.0220	-0.111	-0.0340	1.003*	0.0803
	(0.0800)	(0.0826)	(0.0955)	(0.899)	(0.538)	(0.134)
Observations	290	290	290	290	290	290
Controls	X	X	X	X	X	X

Note: Robust standard errors reported. Regressions are at municipality-year level (metropolitan area-year level for 1993-1998 difference) and include year fixed effects. Sample includes all municipality-years in which CPI price quotes were collected. See Table 1.2 for detailed notes on the instrumental variable construction. All regressions control for lagged variables of log wages, share completing secondary school, share completing university, share of women in workforce, share speaking an indigenous language, share of employment in agriculture, manufacturing, and services.

Table 1.7: Effect of local labor demand shocks on product substitution

	(1) Unweighted	(2) Unweighted	(3) Weighted	(4) Weighted
Shift-share instrument	0.240*	0.152*	0.277**	0.253***
	(0.131)	(0.0847)	(0.127)	(0.0911)
Observations	290	289	290	289
Controls		X		X

Note: Robust standard errors reported. Regressions are at municipality-year level (metropolitan area-year level for 1993-1998 difference) and include year fixed effects. Sample includes all municipality-years in which CPI price quotes were collected. See Table 1.2 for detailed notes about the instrumental variable construction. The dependent variable in columns 1 and 2 is the percentage of items that experienced a substitution. Columns 3 and 4 use a weighted substitution measure that accounts for differences in consumption weights across product categories, defined in Equation 1.5. Columns 2 and 4 control for lagged variables of log wages, share completing secondary school, share completing university, share of women in workforce, share speaking an indigenous language, share of employment in agriculture, manufacturing, and services.

Table 1.8: Quantification results

Panel A: Across Locations

	Consumption shares (mean)		Effect of 10 pp Increase in Local Labor Demand				
	Goods	Services	Mean	SD	p25	p50	p75
	0.463	0.216	-0.017	(0.100)	-0.063	0.0005	0.044

Panel B: Across Income Quintiles

	Consumption shares (mean)		Effect of 10 pp Increase in Local Labor Demand				
Quintiles	Goods	Services	Mean	SD	p25	p50	p75
1	0.514	0.147	-0.154	(0.149)	-0.225	-0.128	-0.074
2	0.533	0.176	-0.134	(0.143)	-0.222	-0.135	-0.059
3	0.512	0.195	-0.088	(0.157)	-0.176	-0.07	-0.012
4	0.483	0.209	-0.031	(0.138)	-0.111	-0.052	0.086
5	0.410	0.253	0.084	(0.132)	-0.00006	0.085	0.181

Note: Panel A shows the average expenditure share on goods, average expenditure share on services, and distribution of the estimated effect of a 10 percentage point increase in the growth rate of local labor demand on local non-housing inflation across 46 metropolitan areas in the analysis sample. Expenditure shares are calculated from the 2000 ENIGH survey. Effects are calculated by multiplying product group and metropolitan area-specific expenditure shares by point estimates from the main analysis. Panel B conducts the same exercise but separately for households in the five income quintiles: quintile 1 consists of the poorest 20% of households in each metropolitan area, and quintile 5 the richest 20% of households in each metropolitan area in the ENIGH survey.

Chapter 2

Estimating the Gender Wage Gap in Kenya

2.1 Introduction

In both developed and developing economies, women on average earn less than men. The presence and persistence of the gender wage gap has been a core observation in labor and development economics. In the US, women working full-time, year-round are paid 84% of what men are paid (U.S. Department of Labor 2024). In the European Union, data from 2022 show that women's gross hourly earnings were on average 12.7% below those of men (Eurostat, 2022). In the developing world, empowering women has been a key policy focus; for example, the United Nation's Millenium Development Goals include one that is to "promote gender equality and empower women". Various interventions have been implemented by governments and policy organizations in the developing world to achieve gender equality in the labor market, including initiatives to increase women's educational attainment, lower barriers to entrepreneurship, improve women's access to jobs and opportunities, and reduce workplace discrimination. Yet surprisingly, we understand little about the magnitude of and contributors to the gender wage gap in low- and middle-income countries (LMICs). Approaches to estimate and account for the gender wage gap typically suffer from the omitted variable bias which violates the unconfoundedness assumption: Many factors potentially affect wages and differ between men and women, but data on these variables are often scarce in LMICs, creating difficulties for rigorous identification.

This present study aims to address this issue by leveraging both highly detailed individual-level human capital and labor market data and a combination of machine learning and econometric methods. On the data front, we use a survey of 5,878 prime-age adult individuals in Kenya, a lower middle income country with a GDP per capita of $2,190. Besides measuring an individual's earnings, labor supply, occupation, education, and demographics information, the survey also includes detailed modules on cognitive abilities, economic preferences (e.g., risk aversion), Big Five personalities, and the type of tasks that individuals do on the job. These questions allow us to measure typically unobserved potential determinants of labor market outcomes and rigorously estimate the gender wage gap. In terms of methods, we combine classic selection-on-observables regressions with machine learning approaches. We use machine learning (ML) methods to first predict wages using a number of covariates to examine the importance of gender as a predictor, and subsequently use Double/Debiased Machine Learning (DML) to allow for nonlinearities in the regression model. Following approaches in the labor literature, we also perform an Oaxaca-Blinder decomposition to decompose the gender wage gap into a component explained by covariates and an unexplained component.

We find that gender is a key predictor of wages: It is the third-most important predictor out of 104 variables in a machine learning model, and removing it as a predictor substantially lowers the predicted wage for men and increases it for women. The average predicted wage becomes 13 log points higher for women and 15 log points higher for men. We estimate that the raw gender wage gap between women and men is 78 log points, or 54 percent. Controlling for education, experience, demographics, and occupation, the gender wage gap becomes 50 log points, or 39 percent. Our rich data on cognition, economic preferences, personality, and job task indices allow us to further account for 18% of the remaining gender wage gap, yielding an estimated difference between men and women of 41 log points, or 34 percent. The Oaxaca-Blinder decomposition shows similar results: Using the full set of controls, 51% (39.4 log points) of the raw gender wage gap can be explained by covariates, leaving 49% (38.6 log points) of the gap unexplained. We also examine the gap in hours worked between men and women. We find that women on average work 12.1 fewer hours per week, and accounting for the full set of controls narrows the gap to 6.8 hours.

We begin by introducing the data. The fifth round of the Kenya Life Panel Survey (hereafter, KLPS-5) was administered in 2023 and surveyed 5,878 individuals who were in Western Kenya and participated in a public school deworming program between 1998 and 2003. Individuals in the sample have an interquartile range of ages 35-39,[1] at the midpoint of their prime working years, reside in various urban and rural areas across Kenya, and are 56% female. We describe the rich variables measuring cognition, job task content, economic preferences, and personality. We then begin the analysis by applying machine learning techniques to predict log wages. We find that gender emerges as one of the most important predictors: It ranks third in its importance among 104 explanatory variables. If we remove gender as a predictor, we would over-predict female wages by 13 log points and under-predict male wages by 15 log points, which demonstrate the important role of gender in determining wages even after considering a rich set of covariates.

We then proceed to regression analyses. The raw gender wage gap is 78 log points. Controlling for educational attainment, experience, and occupation, the gender wage gap narrows to 50 log points. Adding our measures of cognition, task indices, economic preferences, and personality, we find that the gender wage gap further narrows to 41 log points. The additional measures explain 18% of the gender wage gap that remains unexplained by

[1] The 10th percentile is 34 and the 90th percentile is 41.

standard controls, but still leave a substantial portion unexplained. We also examine the gender gap in hours worked and find qualitatively similar results: The additional measures help explain part of the gender gap, but a substantial portion remains. We confirm the robustness of our results by applying Double/Debiased Machine Learning (DML), which is a selection-on-observables method that flexibly controls for nonlinear relationships between the covariates and the outcome or the main explanatory variable of interest. Furthermore, following prior work in labor economics, we perform an Oaxaca-Blinder decomposition. Similar to the regression results, we find that 51% (39.4 log points) of the raw gender wage gap can be explained by the full set of covariates, leaving 49% (38.6 log points) of the gap unexplained. We conduct a series of sensitivity analyses to show robustness of our main findings, and conclude with a discussion of the potential roles of discrimination and social norms in contributing to the gender gap in labor market outcomes.

We aim to contribute to several strands of literature in this study. First, we extend existing work on the gender wage gap. There is a longstanding literature examining the evolution of and contributors to the gender wage gap in the developed world. Recent work from OECD countries shows that even though significant advances have been made to achieve gender equality, the gender wage gap still persists: For example, in 2010, the median women earned 19% less than men in the US, 19% in the UK, 14% in Sweden, 28% in Japan, and 14% in Australia (Kunze, 2018). Another point that this literature makes is that conventional human capital variables (education and labor market experience) together explain little of the gender wage gap in the aggregate. This is in part due to the reversal of gender differences in education, where in countries such as the US women have actually overtaken men in years of education, and in part due to the substantial reduction in the gender experience gap (Goldin, 2014; Blau and Kahn, 2017). Occupational sorting remains a significant cause of the gender wage gap: In the US, the largest identifiable causes of the gender wage gap are differences in the occupations and industries where women and men are most likely to work (US Department of Labor, 2024). However, a substantial portion of the gender wage gap remains unaccounted for even after controlling for occupation. For example, Blau and Kahn (2017) find that in 2010 PSID data, 38 percent of the gender wage gap cannot be explained by education, experience, region, race, unionization, industry, and occupation variables.

In low- and middle-income countries (LMICs), due to the stronger presence of traditional gender norms and the limitations of legal and administrative institutions in addressing gender-based discrimination, one might expect the gender gap to be larger. However, despite sustained academic and policy interest in empowering women in the labor market, while there has been some work examining the gender gap in labor force participation (Verick, 2014; Klasen, 2019), well-identified evidence on the gender wage gap in LMICs remains scarce. Appleton, Hoddinott, and Krishnan, 1999 estimate the gender wage gap in Cote D'Ivoire, Ethiopia, and Uganda, and find that education, experience, and marital status help explain a large part of the wage differences between men and women. Nordman, Robilliard, and Roubaud (2011) estimate the gender earnings gap from seven major West African countries with data from 2001-2003. They find large gender earnings gaps in all the cities in their sample, and that 46-67% of the gender earnings gap cannot be explained by education, experience, and marital status. Nordman, Sarr, and Sharma (2015) estimate the gender wage gap among a sample of formal urban workers in Bangladesh. They measure cognition through literacy and numeracy tests and administer survey questions about workers' personality, and find that the gender wage gap narrows but not completely vanishes when cognitive skills are included. We confirm their findings in a different context with a larger sample of formal and informal workers in Kenya that is representative of the main occupation groups of the country.

Studies on the gender wage gap in Kenya and Sub-Saharan Africa have evolved over time, highlighting key drivers of disparities. Agesa, Agesa, and Dabalen (2013), using 2004–2005 KIHBS data, found that gender differences in characteristics and returns widened wage gaps across distributions, driven by education, occupation, industry, and region. Alarakhia et al., 2023 analyzed 10 countries in the region, reporting that women earn 81 cents per dollar earned by men, with occupational segregation playing a significant role. Van Den Broeck, Kilic, and Pieters, 2023 examined rural and urban areas in Malawi, Tanzania, and Nigeria, finding rural gaps largely explained by worker characteristics, while urban gaps varied more significantly and were less explained by observable factors. The UN Women (2023) study on Kenya revealed an adjusted gender pay gap of 9.5% after accounting for labor market characteristics like education, occupation, marital status, and formality, reducing the raw gap by nearly half but highlighting persistent inequities. Alwago, 2024, using STEP survey data for urban Kenya, attributed 70 to 94. 7% of the wage gap to differences in endowments, but also identified evidence of discrimination and favoritism toward men. While our study aligns with the broader focus of existing research on gender wage disparities, it offers several key contributions that advance the literature. Unlike prior studies, which often rely on administrative data and focus predominantly on urban populations, our analysis incorporates rural Kenya, providing a more comprehensive perspective on the gender wage gap across different settings. Additionally, we employ richer and more nuanced measures of human capital that go beyond traditional metrics such as education and occupation. This allows for a deeper examination of the factors contributing to wage disparities.

Furthermore, this paper contributes to the literature studying the role of cognitive and non-cognitive skills in labor market outcomes. Studies from high-income countries have shown that cognitive skills are rewarded in the labor market (Cawley, Heckman, and Vytlacil, 2001; Heckman, Stixrud, and Urzua, 2006; Lin, Lutter, and Ruhm, 2018; Hermo et al., 2022). Evidence from LMICs also points to a similar direction. Boissiere, Knight, and Sabot (1985) study this with Raven's Matrices test scores and tests of reading achievements and mathematics among a sample of urban wage-labor force in Nairobi, Kenya and Dar es Salaam, Tanzania. They find that the coefficient on cognitive ability is large and statistically significant. Hanushek and Woessmann (2008) review the evidence from Ghana, Kenya, Morocco, South Africa, Tanzania, Pakistan, and conclude that the available estimates of the impact of cognitive skills on outcomes suggest strong economic returns within developing countries. We build on this literature by introducing detailed cognition measures evaluating multiple dimensions of an individual's ability, and confirm prior findings of returns to cognitive abilities in the Kenyan context. A recent literature also highlights the returns to non-cognitive skills (Deming, 2017; Weidmann and Deming, 2021; Edin et al., 2022). Our results that consistently show the importance of engaging in interpersonal tasks on wages affirm these finding in a new context. Furthermore, our job task indices build on the literature applying a task-based approach to human capital accumulation. Starting from Acemoglu and Autor (2011) and Autor and Handel (2013), economists have focused on the role of not only occupations, but also worker's task assignments within occupations. Dicarlo et al. (2016) and Lo Bello, Puerta, and Winkler (2019) construct analytical, interpersonal, and routine/manual indices from World Bank skill surveys across 11 developing countries; we build on their approach by applying the task-based approach to the analysis of wages, as well as by examining gender differences in analytical and interpersonal task indices and their effects on the gender wage gap.

This paper also contributes to a recent literature in labor economics that applies machine learning techniques to study economics questions of interest. Works such as Bonaccolto-Töpfer and Briel (2022), Böheim and Stöllinger (2021), Vafa, Athey, and Blei (2023), and Bach, Chernozhukov, and Spindler (2024) have used causal machine learning methods to estimate the gender wage gap, with varying results on how different the coefficients are between the regression-based methods and ML methods. Most existing papers use post-double LASSO, which applies a two-step lasso regression procedure to select covariates (Belloni, Chernozhukov, and Hansen, 2014). Our application of Double/Debiased Machine Learning (DML) (Chernozhukov, Chetverikov, et al., 2016; Chernozhukov, Nekipelov, et al., 2018), a method that builds on post-double LASSO, enriches this literature by demonstrating its use to study a real-world policy question where a large number of covariates are included.

2.2 Data

The Kenya Life Panel Survey (KLPS) began in 2003 to track a representative sample of approximately 7,500 students enrolled in grades 2 to 7 in schools under the randomized Primary School Deworming Program (PSDP) from 1998 to 2003. The fifth round of KLPS (KLPS-5) was administered in 2023 to observe life outcomes for the same sample of individuals. A notable feature of the KLPS is the commitment to tracking all respondents selected at baseline regardless of whether they have relocated within Kenya or beyond, resulting in high overall effective tracking rates, with 86.5% ever surveyed across all rounds (Walker et al., 2023). This cohort has been followed across multiple survey waves, providing rich data on their health, economic, and cognitive outcomes (Hamory et al., 2021). The KLPS-5 analysis sample includes 5,878 individuals aged 31-47 who are in their prime working years. 44% are male and 56% are female. Additionally, 56% of the sample reside in rural areas, 18% reside in Nairobi, and 26% reside in urban areas outside of Nairobi.

We measure workers' earnings from three sources: agricultural profit, self-employment profit and earnings from wage employment over the past 30 days (12 months for agriculture). We also ask about hours worked on each activity over the past 7 days. We calculate the hourly wage by dividing total weekly earnings by the hours worked, where weekly earnings are the sum of monthly earnings from all three employment sources divided by 30/7=4.29. We exclude individuals who report zero earnings from all three sources or report zero hours worked in the past 7 days. This leaves us with a sample of 5,063 (86.5%) of individuals with a non-zero wage. Using the exchange rate of 1 USD = 140 Kenyan Shillings, an average worker in the KLPS-5 sample works 40.3 hours per week, earns \$89 per month, and has an hourly wage of \$0.72. The median worker works 40 hours per week, earns \$36 per month, and has an hourly wage of \$0.24.[2]

The KLPS-5 survey includes a detailed cognition module to measure individual ability, which typically affects earnings and is correlated with educational outcomes. Cognitive performance was measured using an adapted version of the Harmonized Cognitive Assessment Protocol (HCAP), a neuropsychological assessment battery that comprises tests and items in common with other international partner studies of the US Health and Retirement Study (HRS), which evaluates multiple domains: general cognition, memory, executive functioning, language, orientation, and visuospatial processing. Tests included items related to orientation to place and time; measures of memory including the CERAD 10-Word List Learning, East Boston Memory Test (EBMT), and Logical Memory Test; measures of executive functioning/attention including Raven's Progressive Matrices, Similarities and Differences, Token Test, Go-No-Go, Trail Making, and Symbol Cancellation; measures of language/fluency including animal fluency and various tests of object naming; and measures of visuospatial functioning including CERAD Constructional Praxis, interlocking pentagons, and clock drawing. Details of these cognition tests (and validation for this sample) are described in Gross et al. (2024). In total, 27 sections of questions were administered. Responses were aggregated into a general cognition score and five domain-specific scores (executive functioning, orientation, memory, language, visuospatial) following the Confirmatory Factor Analysis (CFA) approach in Gross et al. (2024), which is a common method in the psychology literature that studies cognition for summarizing cognitive performance measures into a few domain-specific scores that are easily interpretable. Appendix B.1 includes more details.

We also administer a job task content module that asks individuals about the task content of their primary job. These questions follow the approach used in the Skills Toward Employment and Productivity (STEP) surveys commissioned by the World Bank (Dicarlo et al., 2016; Lo Bello, Puerta, and Winkler, 2019; De La Rica, Gortazar, and Lewandowski, 2020) and US-based measures of job task content (Acemoglu and Autor, 2011; Autor and Handel, 2013). These measure three dimensions of the task content of a worker's job: analytical tasks, which measure the degree of analytic reasoning skills used at work; interpersonal tasks, which measure the degree of interactive, communication, and managerial skills used at work; routine/manual tasks, which measure the degree of physical movement and repetitive work. We sum up the scores from questions in each category following the approach taken by Lo Bello, Puerta, and Winkler (2019) and standardize each index. Appendix B.1 includes more details on the specific questions asked and the formula used to construct the three task

[2]Unless noted otherwise, all results in U.S. dollars in this paper use the exchange rate of 1 USD = 140 Kenyan Shillings.

indices. Lastly, to account for potential differences in home care responsibilities between men and women, we include the number of children and the number of older adults over 65 years old in the household in our analyses.[3]

Furthermore, the KLPS-5 survey measures a wide range of typically unobserved variables that could differ between genders and affect individuals' labor market outcomes. We measure the stress level of an individual's primary job following the approach used in the Health and Retirement Surveys (HRS). Respondents were assessed with a 15-item questionnaire taken from the widely used Karasek Job Content Questionnaire (Karasek et al., 1998). They were asked to rate their degree of agreement, on a scale from 1 to 4, with items assessing job demands (e.g., time pressure, amount of work, little job security) and a sense of control or decision-making capacity. We calculate a standardized job stress index from these questions. As individuals' traits could affect their labor market outcomes, we measure individuals' Big Five Personality traits—extroversion, agreeableness, openness, conscientiousness, and neuroticism—from 15 questions commonly used in psychology studies that ask about whether they agree or disagree with statements such as "I tend to feel depressed" and "I assume the best of people" on a 1 to 5 scale. We also ask respondents two additional questions—"Is it impossible to escape a predetermined fate?" and "How satisfied are you with life?"—to measure their attitude toward fatalism and their life satisfaction. We administer an economic preferences module to elicit individuals' attitude toward taking risks, reciprocity, out-group altruism, time preferences, and ambiguity preferences. We include 9 variables that measure economic preferences in our analysis. Appendix B.1 and B.1 describes the construction of these variables in detail.

We record each individual's primary occupation by asking about the most important occupation they have worked in over the past 12 months, and standardize them into 49 occupation codes. We compare the mix of occupations in the KLPS-5 sample with the most recent Kenyan census, Demographic and Health Survey (KDHS), that took place in 2022. In the KDHS data, occupation levels were provided at both the 1-digit and 3-digit levels. Using the ISCO-08 classification, we aggregated these to the 2-digit level, enabling a consistent mapping to KLPS occupations. The initial sample size was 134,555, representing individuals aged 31 to 47—comparable to the KLPS sample age range. After excluding individuals who had not worked in the past 12 months, the final sample was reduced to 88,805. We find that the distribution of occupations at the 1-digit level is similar across both datasets. Appendix Table B.2 shows the result. This indicates that our sample captures a broadly representative snapshot of individuals in the Kenyan labor market.

[3]Specifically, the question asks: How many adults from age 66 (other than yourself)/children under the age of 18 are in your household, "eat from the same pot" and spend 4 nights or more in an average week sleeping in your home?

2.3 Descriptive Facts

Labor market and education outcomes by gender

We first examine gender differences in labor market outcomes. Table 2.1 shows the average outcomes for women and men in the KLPS-5 sample. We see striking differences in labor supply, earnings, and wages. On average, men work 46.6 hours per week, 32% higher than women, who work 35.3 hours on average. Men also earn significantly more than women: On average, men's monthly earnings are 138.3 USD, whereas women's monthly earnings are 49.5 USD. Part of the differences in earnings reflects differences in labor supply, but the hourly wage also differs substantially. On average, men earn 1.04 USD per hour——more than twice as much as women, who earn who earn 0.45 USD per hour on average. Compared to women, men are slightly more likely to engage in self-employment (46.4% vs 40.5%), much more likely to engage in wage employment (52.3% vs 33.2%), and much less likely to experience unemployment over the past 30 days (5.2% vs 13.2%).

One factor that could explain the striking gender differences in labor market outcomes is a difference in educational attainment. However, the gender gap in education is small. Women on average have slightly lower levels of education than men in our sample: The average woman has attained 8.9 years of education, while the average man has attained 9.9 years of education. The difference in education is small, which parallels similar findings in high-income countries where women have closed the education gap with men, sometimes even exceeding men (Goldin, 2014).

Occupational sorting by gender

Another common source of the gender wage gap is occupational sorting. Women may disproportionately work in occupations with lower earning potential due to factors such as preferences and social norms. Table 2.2 shows the top 20 out of 49 occupations, sorted by their employment shares in the full sample for both men and women, and the share of men and women working in these occupations. Women are heavily represented in agriculture and fishing, with 33.2% of them working as farmers compared to 17.1% of men. Retail and commercial activities, such as being a seller or street vendor, are also disproportionately female-dominated, with 17.9% of women and only 5.0% of men engaged in these roles. Similarly, women are much more likely to be employed in domestic work, which accounts for 8.1% of their employment compared to only 0.8% of men's. Some of these occupations are often characterized by lower earning potential, reinforcing income disparities between the genders. In contrast, men dominate several occupations that tend to have higher physical demands or are socially perceived as male-oriented. For instance, 7.7% of men work in fishing, compared to only 0.6% of women. Similarly, bicycle and motorbike taxi work is almost exclusively male, with 6.0% of men participating and no women recorded in this occupation. Construction-related jobs also exhibit stark gender imbalances, with roles such as mason (4.9% for men versus 0.1% for women) and unskilled construction laborer (4.0% for men versus 0.3% for women) heavily skewed toward men.

Gender differences in cognition

Are there substantial gender differences in cognition? Figure 2.1 shows the average differences, in standard deviations, between men and women for the general cognition score as well as the five component-specific cognition scores. Men outperform women on the general cognition score by 0.21 standard deviations, with notable variation across specific cognitive skills. The largest disparities are observed in executive function and visuospatial skills, where men score 0.31 and 0.28 standard deviations higher, respectively. Spatial orientation and language show smaller gaps, with men scoring 0.06 and 0.10 standard deviations higher, while memory exhibits the smallest difference at just 0.04 standard deviations. These findings suggest that gender differences in cognitive performance are persistent, though not uniform, across domains. The rightmost column in Figure 2.1 shows the differences in the lagged Raven's Matrices test score, which was administered 11 to 13 years before the KLPS-5 data were collected, among 5,264 out of 5,878 individuals who were in both the KLPS-3 and KLPS-5 samples. This measure reveals the largest gender disparity, with men outperforming women by 0.41 standard deviations. As the Raven's Matrices test corresponds to components of the executive functioning abilities, the observed difference here is consistent with the large difference in executive functioning score observed in the current round of KLPS.

Gender differences in job task contents

Previous discussions have shown that men and women tend to work in different occupations. On top of that, the contents of the jobs that men and women engage in can also differ significantly, even for the same occupation. Figure 2.2 shows the average differences between men and women in the analytical, interpersonal, and routine/manual task indices. Compared to the average job performed by women, the average job performed by men scores 0.52 SD higher on the analytical task index, 0.67 SD higher on the interpersonal task index, and 0.57 SD higher on the routine/manual task index. The tasks performed by men tend to be high in abstract thinking, more intensive in interacting with others, and feature a larger share of repetitive and manual work. Appendix Table B.1 shows the gender differences in each of the task content measures that are inputs to the indices. The difference in the analytical task index stems from substantial disparities in tasks such as reading long documents (0.39 SD), performing math at work (0.40 SD), and engaging in activities that involve learning new things (0.33 SD). Among activities that demonstrate and develop interpersonal skills, men are much more likely to supervise others (0.48 SD), collaborate with coworkers (0.65 SD), and make presentations at work (0.43 SD). For the components of the routine/manual task index, while men and women are equally likely to perform repetitive tasks at work, men are more likely to operate heavy machinery at work (0.43 SD), repair or maintain equipment at work (0.27 SD), and find their work physically demanding (0.18 SD).

These results potentially reflect differences in physical strength and endurance between men and women.

Figure 2.6 presents a Radar chart that summarizes the gender differences across five skills and human capital categories—educational attainment, cognition, analytical task index, interpersonal task index, and routine/manual task index—–for females and males. The figure displays, in standard deviations, the average male and female outcomes across the above mentioned categories, allowing for easy comparison across domains of skills and human capital measures. Across all categories, females have negative scores while males have positive ones, indicating that males consistently outperform females in these dimensions. Among the categories, the largest gender gap is observed in interpersonal skills, where the difference between males and females is 0.67 SD. On the other hand, the smallest gap appears in Cognition, where the difference is 0.21 SD. This indicates that gender differences in cognitive abilities are less pronounced compared to job skill measures and educational attainment.

Gender differences in other observables

We close this section with a discussion on the gender differences in other measures that could affect wages: economic preferences, personalities, and job stress. Personality traits have the potential to explain individual labor market outcomes; studies in the psychology literature tend to find a positive association between earnings and the traits of openness, conscientiousness, and extraversion, and a negative association between earnings and the traits of agreeableness and neuroticism (Alderotti, Rapallini, and Traverso, 2023). In light of those findings, we examine gender difference in Big Five personalities, shown in Figure 2.3 in standard deviations. Men are more likely than women to demonstrate extraversion, agreeableness, conscientiousness, and openness, while women are more likely to than men to demonstrate neuroticism. Among economic preferences measures, women are more likely than men to display altruistic tendencies, both for members of the in-group and out-group.[4] Women are 0.78 SD more likely than men to believe that it is impossible to escape a predetermined fate, when presented with that statement. There are no statistically significant gender differences on other economic preference variables, such as ambiguity, risk aversion, reciprocity, and patience. We do not find gender differences in life satisfaction (diff = 0.002 SD). We also do not find gender differences in the stress level experienced on the job between men and women (diff = 0.01 SD) or the number of older adults over 65 years old in the household (diff = 0.00). However, women on average have 0.34 more children under 18 years old than men in their household (p-value < 0.001). Figure 2.4 illustrates these results. Even though the differences in the number of children is not causal and likely include differences in household characteristics between women and men in the sample, such difference may point to a higher burden for women in the sample to engage in home production. In subsequent

[4]We ask respondents how much they would donate to charity if they unexpectedly received 3,200 shillings ($22.9) today. We ask the same question again, but specifying that the respondent knew that it was a charity helping people in their ancestral home area, and again specifying that the respondent knew that it was a charity helping people in other parts of Kenya other than their ancestral home area.

empirical analyses, we control for these variables along with the others mentioned in this section.

2.4 Machine Learning Prediction of Wages

Does an individual's gender predict their wages? In this section, we document the role of gender and other observables in predicting wages through a series of machine learning exercises. We use two sets of explanatory variables (i.e., features): the full set of 104 variables that contain education, experience, demographics, occupation, cognition, task indices, personality, preferences, home care, and job stress measures, and a "basic" set of 55 education, experience, demographics, and occupation variables.[5] All exercises use the log wage as the outcome and 5-fold cross-validation with a random forest regressor, which allows features to affect the outcome in a non-linear manner and avoids overfitting by creating random subsets of the features and building smaller trees using those subsets. The results are robust to using other machine learning algorithms, such as LASSO and XGBoost.

We first predict the log wage for the entire sample of individuals. The random forest algorithm achieves an average test R-squared of 0.3 when we use all features, and an R-squared of 0.28 when we use only the basic set of features. Without any covariate adjustment, women earn 80 log points less than men, which represents a 55% difference. On average, the predicted log wage for women is 69 log points less than men, which represents a 50% difference. If we remove gender as a predictor, the R-squared becomes 0.28 with all other features included. While the R-squared only declines modestly, the predicted gender wage gap shrinks substantially if we do not include gender as a predictor: The predicted gender wage gap becomes 41 log points, which is equivalent to women earning 33% less than men. Importantly, removing gender as a predictor substantially lowers the predicted wage for men and increases the predicted wage for women. The average predicted wage becomes 13 log points higher for women and 15 log points lower for men. This highlights the importance of gender as a predictor of wages: even though on average a model without accounting for gender still maintains predictive power, it cannot accurately predict gender-specific wages. On the other hand, if gender does not matter for wages conditional on other observables, a model without gender as a predictor should not produce drastically different results for the average male and female wages, compared to one with gender as a predictor.

Even though random forest models lack the interpretable coefficients of regression models, which allow us to easily observe which explanatory variables affect the outcomes the most, SHAP (SHapley Additive exPlanations) provides a way to examine the influence of each feature on the model's predictions. It builds on an approach from coalitional game theory that measures each player's contribution to the final outcome (Molnar, 2023). SHAP values decompose a prediction into the contributions of each feature for that observation, which

[5] Following standard approaches in the labor literature, years of experience is calculated by (age - years of education - 6) (Card et al., 2018). We include both years of experience and its squared term.

are analogous to marginal effects in regression but account for feature interactions and non-linearities inherent in random forests. SHAP has been widely used in machine learning to increase the transparency and interpretability of otherwise complex, non-linear models.

Figure 2.5 shows the SHAP beeswarm plot for the features with the highest predictive power in predicting log wages. The plot visualizes the relative importance of features in predicting log wages, along with the direction and magnitude of their effects. Features are ranked by their average SHAP value magnitude, with the most important predictors at the top. Residing in an urban area (urban_binary) is the most important predictor, with higher SHAP values (pink points) for individuals in urban areas pushing predicted wages upward, while lower values (blue points) for rural residents decrease predictions. Following urban residency, gender emerges as the second most important predictor. Its SHAP values reveal a consistent negative impact of being female (blue points) on wages, as the majority of points cluster on the left side of the plot. In contrast, being male (pink points) has a positive effect, pushing predictions toward higher wages. This highlights the substantial role of gender in wage disparities, even after accounting for occupation, education, job characteristics, and individual traits. Other influential predictors include the interpersonal task index, educational attainment, analytical index, executive functioning score, and general cognition score. Consistent with out intuition, all of those predictors have a positive effect on predicted log wages. As a robustness exercise, Appendix Figure B.1 shows the SHAP beeswarm plot from a model with only the education, experience, demographics, and occupation variables as predictors. Gender still emerges as the second most important predictor of wages.

We also perform the same exercises to predict hours worked. Recall from Table 2.1 that men on average work 46.6 hours per week, but women work only 35.3 hours per week. Using the full set of 104 covariates to predict hours worked, the random forest model achieves an R-squared of 0.22. If we remove gender as a predictor, the R-squared becomes 0.21 with all other features included. Without gender as a predictor, the model cannot tailor its prediction to males and females quite well: The average predicted hours worked becomes 1.26 hours higher for women and 1.01 hours lower for men. Appendix Figure B.2 shows the SHAP beeswarm plot that summarizes the contribution of top features to the prediction of hours worked. Gender emerges as the third most important feature, behind the interpersonal index and self-employment (self-employed workers work substantially more hours). Appendix Figure B.3 shows the SHAP beeswarm plot from a model that uses only education, experience, demographics, and occupation as features. Again, gender emerges as the second most important feature, behind only self-employment.

2.5 Regression Analyses

Main results: wages

Having established that many individual characteristics both differ by gender and affect wages, we now use a regression framework to study how the observed gender wage gap once we account for different sets of observable characteristics. We implement the following regression model:

$$\ln Y_i = \beta_0 + \beta_1 Female_i + \beta_2 Sch_i + \beta_3 Exp_i + \beta_4 Exp_i^2 + \beta_5 X_i + \gamma_i + \epsilon_i \qquad (2.1)$$

Here, Y_i is the log hourly wage of individual i, Sch_i is the years of schooling completed, Exp_i is the years of experience, γ_i includes 49 occupation indicators, and X_i is a vector of person-level controls that include different sets of covariates depending on the specification. Table 2.3 shows the results. Column 1 shows that the raw, unadjusted gender wage gap is 78 log points, which is equivalent to a 54 percent difference. Column 2 adds controls for years of educational attainment, experience, location, and any types of agricultural work or self-employment. These are standard control variables that, due to lack of available data, studies examining the gender wage gap in a developing context typically include. The inclusion of these variables reduces the gender wage gap to 62 log points, suggesting that some of the gap is explained by observable demographic characteristics. Urban residence has a positive and significant association with log wages (coefficient = 0.62) and being in Nairobi also has a compounding effect, both of which highlight the wage premium associated with urban locations. Column 3 adds 49 occupation fixed effects but without the controls in Column 2. Column 4 includes both occupation fixed effects and controls in Column 2. The coefficient in Column 4 reduces to 50 log points, suggesting that occupation sorting contributes to the gender wage gap. Column 5 introduces controls for cognitive ability, the number of children, and the number of older adults in the household. The gender wage gap narrows slightly to 48 log points. The general cognition score is positively associated with wages: a one standard deviation increase in the general cognition score increases wages by 19 log points. Once the general cognition score is included, the coefficient on education drops by half and becomes no longer statistically significant, which is unsurprising since cognition and educational attainment are positively correlated. Having more children in the household is associated with a small but statistically significant reduction in wages with a coefficient of 5 log points, which potentially points to a childcare penalty on wages.

Column 6 further introduces the three task indices—abstract, interpersonal, and routine/manual—as controls to account for task differences within the same occupation. There is a positive wage premium of 10 log points associated with analytical tasks and a 12 log point premium associated with interpersonal tasks. These suggest that within the same occupations, workers whose on-the-job activities involve more abstract thinking and interacting with others earn higher wages. The gender wage gap falls further to 41 log points, suggesting that part of the reason why men earn more than women can be explained by the differences in the tasks that men and women perform. As shown in Figure 2.2, women tend to work in jobs that involve a lower degree of analytical and interpersonal tasks. In Column 7, we include the rest of observables that could potentially affect wages: the job stress index, economic preferences and personality variables. We also include the five component-specific cognition scores along with the general cognition score: executive functioning, visuospatial, language, memory, and orientation, as well as the lagged Raven's score from KLPS-3. The coefficient on general cognition score becomes significantly smaller and not statistically significant, as the addition of component-specific cognition scores absorbs some of the effects loaded on the general cognition score in previous columns. The gender wage gap remains at 41 log points, with little movement. Interestingly, high-stress jobs are associated with lower wages, suggesting a positive correlation between wages and amenities. This kind of positive correlation has also been observed by studies on the US labor market, such as Lamadon et al. (2024) and Sockin (2021). Comparing Column 7 with Column 4, we see that cognition, task indices, personality, and preferences explain 18% of the gender wage gap that was left unexplained by occupation, education and demographic variables. This shows that even after accounting for a comprehensive set of potential confounders, there still remains a large portion of the gender wage gap unexplained.

Main results: hours worked

In our sample, women also work fewer hours than men. How much of the gender difference in labor supply can be explained by observable factors? As increasing women's participation in the labor market has been a key policy focus of many governments and organizations, examining the gender gap in hours worked is of interest in its own right. Table 2.4 shows the results from implementing the regression model as specified in Equation (2.1), but changing the dependent variable to weekly hours worked. All individuals are included (except for 1 observation missing gender and 102 observations missing urban/rural status); individuals who do not report working any hours in agriculture, self-employment, and wage employment are coded as working zero hours. Column 1 shows that without any adjustment, women work 12.1 fewer hours per week than men. Column 2 shows that including demographics and education controls narrows the gap to 10.59 hours. As one may expect, higher-educated individuals work more, and engaging in self-employment significantly increases one's hours worked. Column 4 shows that including occupation fixed effects together with demographics and education controls, women on average work 8.73 fewer hours than men. Column 5 introduces additional controls for cognition, the number of children, and the number of older adults in the household, which reduce the gender gap to 8.58 hours. Notably, these additional controls do not significantly affect the coefficients of the other variables, and none of the household composition variables are statistically significant. Column 6 adds controls for three task indices. Analytical and interpersonal task indices are positively associated with hours worked (coefficients of 1.36 and 3.30, respectively), suggesting that workers performing more cognitively demanding or interpersonal tasks tend to work longer hours. This inclusion reduces the gender gap further to 6.98 hours. Column 7 includes personality, economic preferences, and job stress as additional controls, along with component-specific cognition scores and the lagged Raven's score. The gender gap decreases slightly to 6.83 hours, showing that these additional factors explain a small portion of the remaining gap. Comparing Column 7 with Column 4, we observe that the addition of this set of control variables explains approximately 22% of the gender gap in hours worked that remains unexplained by occupation, education, and demographic variables.

Double/Debiased Machine Learning results

As the set of variables that could be correlated with wages and gender is very large—104 to be precise—and many of them are continuous, imposing linearity may bias our estimates. To account for potential non-linearities, we used a recently developed method, Double/Debiased Machine Learning (DML), from the causal machine learning literature (Chernozhukov, Chetverikov, et al., 2016; Chernozhukov, Nekipelov, et al., 2018). This method is useful in a context where many factors may affect both the treatment (here, being a woman) and the outcome. DML builds on previous causal machine learning methods, such as post-double LASSO, which have been applied by several labor studies examining the gender wage gap in high-income countries (Bonaccolto-Töpfer and Briel, 2022; Böheim and Stöllinger, 2021; Bach, Chernozhukov, and Spindler, 2024). As the name suggests, DML reduces the problem to first estimating two predictive tasks: predicting the outcome from the controls, and predicting the treatment from the controls. Then DML combines these two predictive models in a final stage estimation. The approach allows for arbitrary machine learning algorithms to be used for the two predictive tasks, while maintaining many favorable statistical properties related to the final model, such as small mean squared error and asymptotic normality (Econ ML User Guide 2023).

Formally, DML makes the following assumptions about the data generating process:

$$
\begin{aligned}
&Y = \theta(X) \cdot T + g(X, W) + \epsilon \\
&T = f(X, W) + \eta \\
&\quad \mathbb{E}[\epsilon|X, W] = 0,\ \mathbb{E}[\eta|X, W] = 0,\ \mathbb{E}[\eta \cdot \epsilon|X, W] = 0
\end{aligned}
$$

Here Y is the outcome, T is the treatment (being female), X is a set of covariates across which there are potentially heterogeneous treatment effects, and W are potential confounders. What is particularly attractive about DML is that it makes no further structural assumptions on g and f and estimates them non-parametrically using arbitrary non-parametric Machine Learning methods (Econ ML User Guide 2023). Specifically, we rewrite the structural equation as

$$
Y - \mathbb{E}[Y|X, W] = \theta(X) \cdot (T - \mathbb{E}[T|X, W]) + \epsilon \tag{2.2}
$$

and estimate the following two conditional expectation functions with non-parametric machine learning algorithms

$$
\begin{aligned}
q(X, W) &= \mathbb{E}[Y|X, W] \\
f(X, W) &= \mathbb{E}[T|X, W]
\end{aligned}
$$

and compute the residuals

$$
\begin{aligned}
\tilde{Y} &= Y - q(X, W) \\
\tilde{T} &= T - f(X, W) = \eta
\end{aligned}
$$

These are related by the equation

$$\tilde{Y} = \theta(X) \cdot \tilde{T} + \epsilon$$

Since $\mathbb{E}[\epsilon \cdot \eta | X] = 0$, we can run a linear regression to estimate $\theta(X)$. In the implementation, we use a random forest regressor for the first-stage model predicting the outcome and a random forest classifier for the first-stage model predicting the treatment (i.e., being female). We start by imposing homogeneous treatment effects, i.e., $\theta(X) = \theta$, but in the next section we include interaction terms to study how the gender wage gap varies across various segments of the sample.

Table 2.5 shows the results. The top half of the table illustrates the gender gap in log wages and weekly hours worked using DML with both the basic set of education, experience demographics, and occupation controls (Columns 1 and 3) and a full set of controls that include all observables mentioned in Section 2 (Columns 2 and 4). The gender wage gap remains substantial and similar to those predicted by the linear regression: accounting for the basic controls, the gender wage gap estimated by DML is 57 log points, versus 50 log points estimated by linear regression. Accounting for the full set of covariates, the gender wage gap estimated by DML is 41 log points, versus 40 log points estimated by linear regression. Cognition, task indices, personality, and preferences explain 28% of the gender wage gap that was left unexplained by occupation, education and demographics variables. Because our sample is slightly more female (56%) than male (44%), Columns 3 and 4 account for class imbalance by under-sampling women to achieve gender balance in the analysis sample. This is done to improve the accuracy of the final-stage DML estimates, as the first-stage model predicting being female could tend to over-predict female because it is the more common class. The results are similar to the ones generated from the full sample.

The bottom half of the table shows the DML results in a model where the outcome variable is weekly hours worked. The gender gap in hours worked is 7.07 hours with the basic set of covariates and 4.43 hours with the full set of covariates, which are smaller than the 8.73 and 6.69 hours estimated from linear regression, suggesting that non-linearities likely affect the estimation of the gender gap in hours worked more than the gender wage gap. Columns 3 and 4 perform the same under-sampling exercise to account for class imbalance; the coefficients change slightly but are similar to the ones obtained from the full sample. Taken together, the DML results provide us with more confidence that (1) cognition, job tasks, personality, and economic preference measures help explain part of the gender gap in wages and hours worked, and (2) there remain substantial differences between men and women in their wages and hours worked even after accounting for the full set of covariates observed in KLPS-5.

Oaxaca-Blinder Decomposition

Through regression and DML methods, we have examined how observable factors can explain part of the gender gap in wages and hours worked. Another approach commonly applied in labor economics is the Oaxaca-Blinder decomposition, in which we estimate the wage equations separately for men and women and decompose the raw or unadjusted gender gap into one part explained by observed variables and another unexplained part, following Blinder (1973) and Oaxaca (1973). Specifically, we run the following regressions separately for groups $g = \{m, f\}$, where Y_g denotes log wages and X_g denotes observed covariates:

$$Y_m = X_m B_m + u_m$$
$$Y_f = X_f B_f + u_f$$

where X_g is a vector of coefficients. Let b_m and b_f be respectively the OLS estimates of B_m and B_f, and denote mean values with a bar. Then, since OLS with a constant term produces residuals with a zero mean, we have:

$$\overline{Y}_m - \overline{Y}_f = b_m\overline{X}_m - b_f\overline{X_f} = \underbrace{b_m(\overline{X}_m - \overline{X}_f)}_{\text{Explained Part}} + \underbrace{\overline{X}_f(b_m - b_f)}_{\text{Unexplained Part}} \tag{2.3}$$

The first term on the far right-hand side of Equation (2.3) is the impact of gender differences in the explanatory variables evaluated using the male coefficients. The second term is the unexplained differential and corresponds to the average female residual from the male wage equation. This residual corresponds to an experiment where we compare a woman's actual wage with her predicted wage from the male equation, and one can think of the predicted wage as the outcome of a discrimination case in which a firm that previously was found to have discriminated against women is now required to treat women the same as it treats men Blau and Kahn (2017). The decomposition is typically performed using men as the reference group, but we perform the decomposition using women as the reference group as well in the appendix. Compared to the pooled-sample approach that we carried out with regressions and DML, Oaxaca-Blinder decomposition structurally assumes that observables affect male and female wages differently, whereas the pooled-sample approach can be considered in a potential outcomes framework where one's gender is as good as randomly assigned and observables should not impact one's labor market outcome differently by gender. As a result, it is expected that the output from the two approaches will differ.

Table 2.6 shows the decomposition results for the gender wage gap. The part that is explained by the basic set of covariates (education, experience, demographics, occupation) is 30.7 log points, or 39% of the raw gender wage gap; the unexplained part is 47.3 log points, or 61% of the raw gender wage gap. Using the full set of controls, 51% (39.4 log points) of the raw gender wage gap can be explained by covariates, leaving 49% (38.6 log points) of the gap unexplained. These estimates are very close to the pooled-sample regression results that show a 50 and 41 log point gender wage gap unexplained by the basic and full set of controls, as shown in Table 2.3. Table 2.7 shows the decomposition results for the gender gap in hours worked. The part of the gap explained by the basic set of covariates (education, experience, demographics, occupation) is 1.35 hours, or 11% of the raw gender hours gap; the unexplained part is 10.75 hours, or 89% of the raw gap. Using the full set of controls, 30% (3.64 hours) of the raw gender hours gap can be explained by covariates, leaving 70% (8.46 hours) of the gap unexplained. These estimates are slightly different from the ones obtained from pooled-sample regressions shown in Table 2.4, but still within the same range and similar. To complete the analysis, Appendix Tables B.3 and B.4 show the Oaxaca-Blinder decomposition results using female as the reference group, which differ from those presented above as it is known that estimated components may change depending on the choice of the reference group (Oaxaca and Ransom, 1994).

Interaction terms

How does the gender gap vary across locations, education levels, cognitive abilities, and job task contents? In this section we examine the heterogeneity by including interaction terms with the gender dummy. Specifically, we interact the gender dummy with the following variables along which the gender wage gap could differ: urban/rural residency, living in Nairobi, any agricultural work, any self-employment, educational attainment, cognition (general cognition score, each one of the five component-specific cognition scores, lagged Raven's score), experience, experience squared, analytical task index, interpersonal task index, and routine/manual task index. Table 2.8 shows the results from both the linear regression model (Columns 1 and 2) and the DML model (Columns 3 and 4). Columns 1 and 3 show the interaction terms with the basic set of controls, and columns 2 and 4 show those with the full set of controls. Although many of the interaction terms are not statistically significant, some patterns emerge. Most notably, in both the linear regression and the DML models, the interaction term on the analytical index is negative (-0.14 and -0.15), whereas the interaction term on the interpersonal index is positive (0.18 and 0.19). These magnitudes are practically significant, given that the gender wage gap (adjusted for all covariates) is 0.41. These results may indicate that the gender wage gap is larger among individuals involved in high analytical tasks, but smaller among individuals involved in high interpersonal tasks. One possible explanation is that gender norms are more entrenched in jobs that primarily involve abstract thinking and mathematical skills, but less so in jobs that primarily involve interpersonal relationships and interacting with customers. To address the concern that there may be limited common support between male and female on these task indices, thereby making the interaction terms hard to interpret, in Appendfix Figure B.4 we display the kernal density plots of each one of the three task indices by gender. There is substantial overlap between men and women. Interestingly, the density graph for the interpersonal index shows evidence of bunching for both male and female, but the density graphs for the analytical and routine/manual indices shoe less signs of bunching and are closer to bell-curve shapes.

The gender wage gap is larger among individuals who engage in self-employment, which echoes findings in the literature that female-run microenterprises tend to underperform their male-run counterparts on many indicators in low- and middle-income countries (Jayachandran, 2021). The gender gap is larger for workers in urban areas, though residing in Nairobi compensates for part of the additional gap. The interaction terms with cognition variables show mixed results: the coefficients on general and component-specific cognition scores are generally insignificant, except for language and visuospatial skills, which exhibit a small but significant negative interaction in some models. Due to their being borderline significant, we do not aim to draw definitive conclusions from them.

Furthermore, we also examine how the gender gap in labor supply varies across demographics and human capital measures. Appendix Table B.5 presents the interaction terms in regression and DML models following the same specification as those implemented for the study of the wage gap. Most coefficients are either close to zero or noisily estimated; however, one observation that stands out is that among individuals whose job involves a high degree of analytical tasks, the gap between women and men in hours worked is smaller. A 1 SD increase in the analytical task index is associated with a 3.15-3.39 smaller gender gap in hours worked, which is a significant magnitude given that the gender gap in hours worked (adjusted for all covariates) is 6.83 hours. This potentially suggests that among high-skilled individuals, women work similar hours to men, though the gender wage gap is larger.

Robustness of results

We conduct a few robustness exercises. First, in Table 2.9 we only use wage from employment as the dependent variable, which excludes agricultural and self-employment income. The gender wage gap remains substantial: on average, among people who engage in wage employment, women earn 41 log points less than men, 34 log points less after adjusting for basic controls, and 30 log points less after adjusting for the full set of controls. The sample is substantially smaller—consisting of 2,260 individuals—because the majority of individuals do not engage in wage employment. Cognition, task indices, personality, and preferences explain about 12% of the gender wage gap that is unexplained by education, demographics, and occupation variables. In Table 2.10, we use log monthly earnings as the dependent variable. The gender gap remains persistent across specifications: on average, in terms of total earnings, women earn 111 log points less than men, 61 log points less after adjusting for basic controls, and 60 log points less after adjusting for the full set of controls. Interestingly, the coefficients on the female dummy in Columns 4 and 7 are almost identical, suggesting that cognition, task indices, personality, and preferences do not help further explain the gender gap in earnings that is unexplained by education, demographics, and occupation variables. In Appendix Table B.6, we conduct the same regression on hourly log wage as in Table 2.3, but winsorize the dependent variable at the 95th percentile, which is 5.75 (equivalent to an hourly wage of 313.42 Kenyan Shillings or 2.24 USD), to examine the impact of outlier responses. In Appendix Table B.7, we conduct the same exercise for hours worked, winsorizing the dependent variable at the 95th percentile, which is 89 hours. For both the wage and hours regressions, the coefficients are very similar in magnitude to those obtained from the full sample, which should increase our confidence in the results. Another concern is that the gender gap in wages could be explained by the gender gap in labor supply—the fact that women work fewer hours might lead to women having less time to build up skills on the job, or supervisors might not appreciate female workers who work part time as much as male workers who work full time. Appendix Table B.8 examines this by including weekly hours worked as an additional covariate in both the regression and DML models. We find that the gender wage gap adjusted for the basic and full set of controls are similar to those in the main specification: 59 vs 50 log points with the basic set of controls and 48 vs 41 log points with the full set of controls, for example. The additional control variables explain 19% of the gender wage gap unexplained by education, experience, demographics, and occupation, similar to the 18% from the main specification.

2.6 Discussion and Conclusion

We have shown that the gender wage gap remains persistent across specifications. The inclusion of cognition, job task context, economic preferences and personalities explain 18% of the gender wage gap left unexplained by education, experience, demographics, and occupation, but it still leaves a large gap of 41 log points. This points to the potential role of other factors in affecting the labor market outcomes of men and women differently. Recent studies in development economics have examined the role of social norms and discrimination (e.g., Bursztyn, González, and Yanagizawa-Drott 2020; McKelway 2019; Buchmann, Meyer, and Sullivan 2024; see Jayachandran (2021) for a review), safety (Siddique, 2022; Becerra and Guerra, 2023), commuting (Cheema et al., 2023; Chen et al., 2024) and work location (Jalota and Ho, 2024) in affecting female labor supply. It is possible that some of those factors also contribute to lower wages earned by women relative to men.

The KLPS-5 survey does not include a detailed module on discrimination or social norms, but we nonetheless asked a few questions related to discrimination experienced by respondents. Table 2.11 shows how individuals' experience of discrimination differs by gender. We ask individuals if they are treated with less courtesy than other people are, if they are treated with less respect than other people are, if they receive poorer service than other people at restaurants or stores, if people act as if they think the respondent is not smart, if people act as if they are afraid of the respondent, and if they are threatened or harassed. Each variable is coded as 1 if the respondent reports that the experience has happened to them more than a few times a year or more (almost every day, at least once a week, a few times a month), and 0 if they report less than once a year or never. Panel A shows both a difference in means between men and women (third column) and a full regression model including all covariates used in the wage analysis (fourth column). Across all variables, we do not discern any patterns that suggest women are more likely to report having experienced discrimination. Perhaps puzzlingly, women are less likely than men to say that they receive poorer service at restaurants or stores or being threatened or harassed. We also ask respondents about the main reason they believe these experiences of discrimination occurred; Panel B shows the share of men and women who cite each reason as the main one. "My education or income level" emerges as the most cited reason, with 48% of men and 49% of women who experienced discrimination choosing that option; social group or family background comes next with 18% of men and 19% of women choosing that option. 12% of men and 11% of women believe aspects of their physical appearance is the main reason for discrimination. Only 1.78% of women and less than 1% of men chose gender as the main reason for discrimination. If we include other reasons that respondents offered, only 2.73% of women and less than 1% of men cite their gender as a reason, suggesting that even among women, gender is not a factor they commonly report as contributing to their experiences of discrimination. These results could reflect differences in self-reporting rather than differences in experiencing discrimination, but further analysis is needed to examine how the experiences of discrimination vary by gender, and whether that affects the labor market outcomes of women and men differently.

Taken together, these results suggest a real puzzle: the gender gaps in wages and labor supply are persistent, and yet our rich set of covariates can only explain about half of it. Nor do women report a higher frequency of discrimination or gender-based discrimination. We believe future research is needed to examine precisely the role of discrimination and social norms in shaping the gender gap in labor market outcomes.

2.7 Figures and Tables

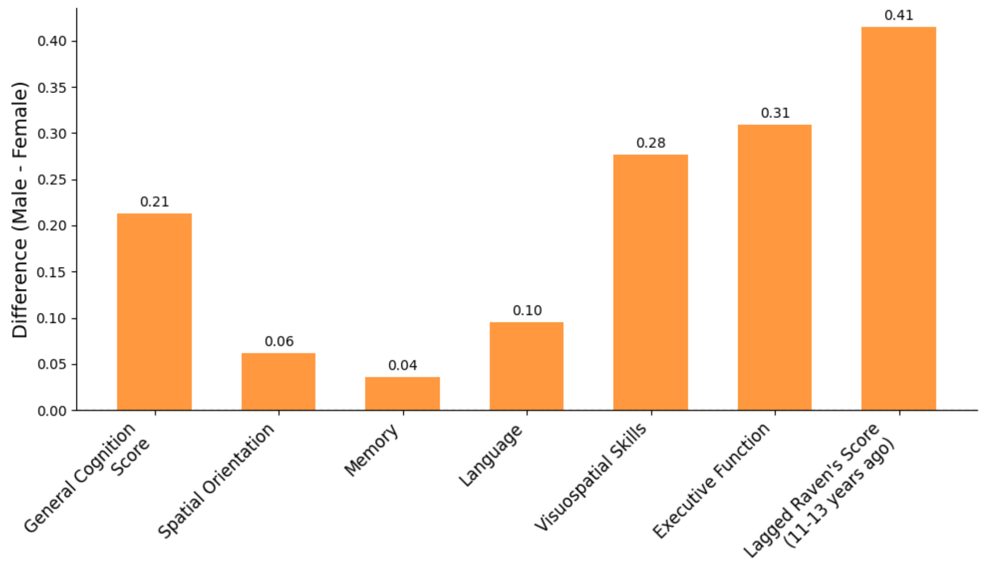


Figure 2.1: Gender differences in cognition

Note: Figure shows the differences in standard deviations between average male and female cognition scores. General cognition score and the five component scores are obtained from the cognition module in KLPS-5. Lagged Raven's scores are obtained from KLPS-3.

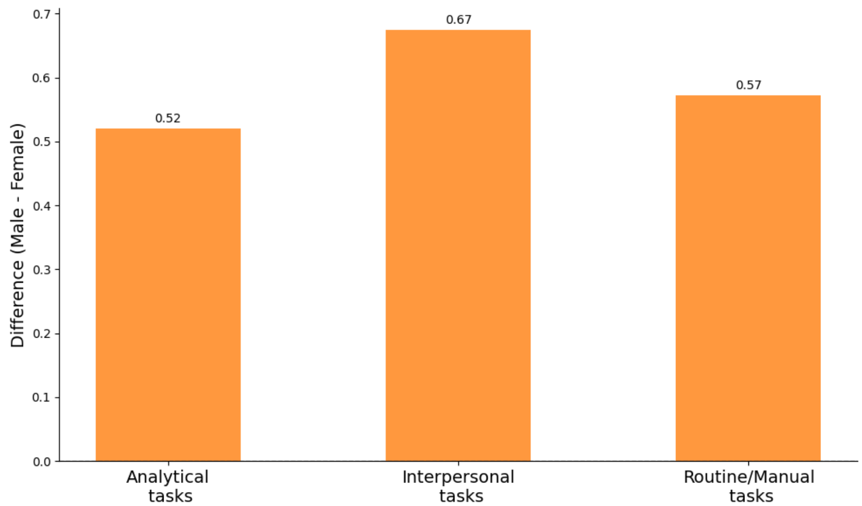


Figure 2.2: Gender differences in task indices

Note: Figure shows the differences in standard deviations between average male and female task index scores.

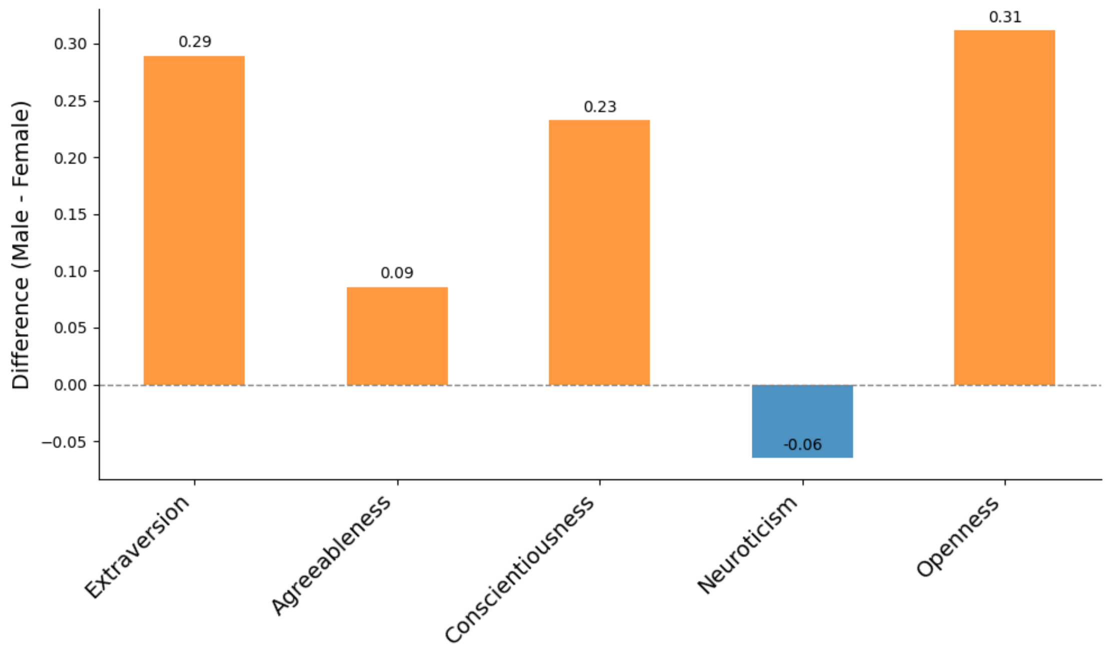


Figure 2.3: Gender differences in Big Five personalities

Note: Figure shows the differences in standard deviations between average male and female Big Five personality indices.

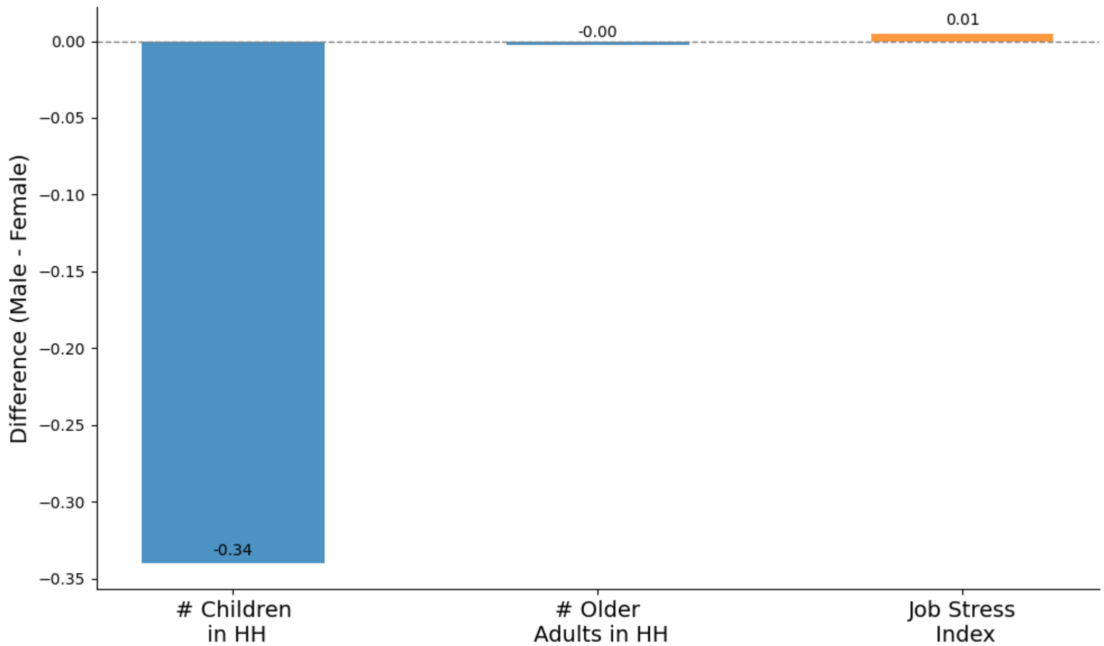


Figure 2.4: Gender differences in home care and job stress

Note: Figure shows the differences in standard deviations between average male and female on: the number of children in the household, the number of older adults over 65 years ago in the household, and the job stress index.

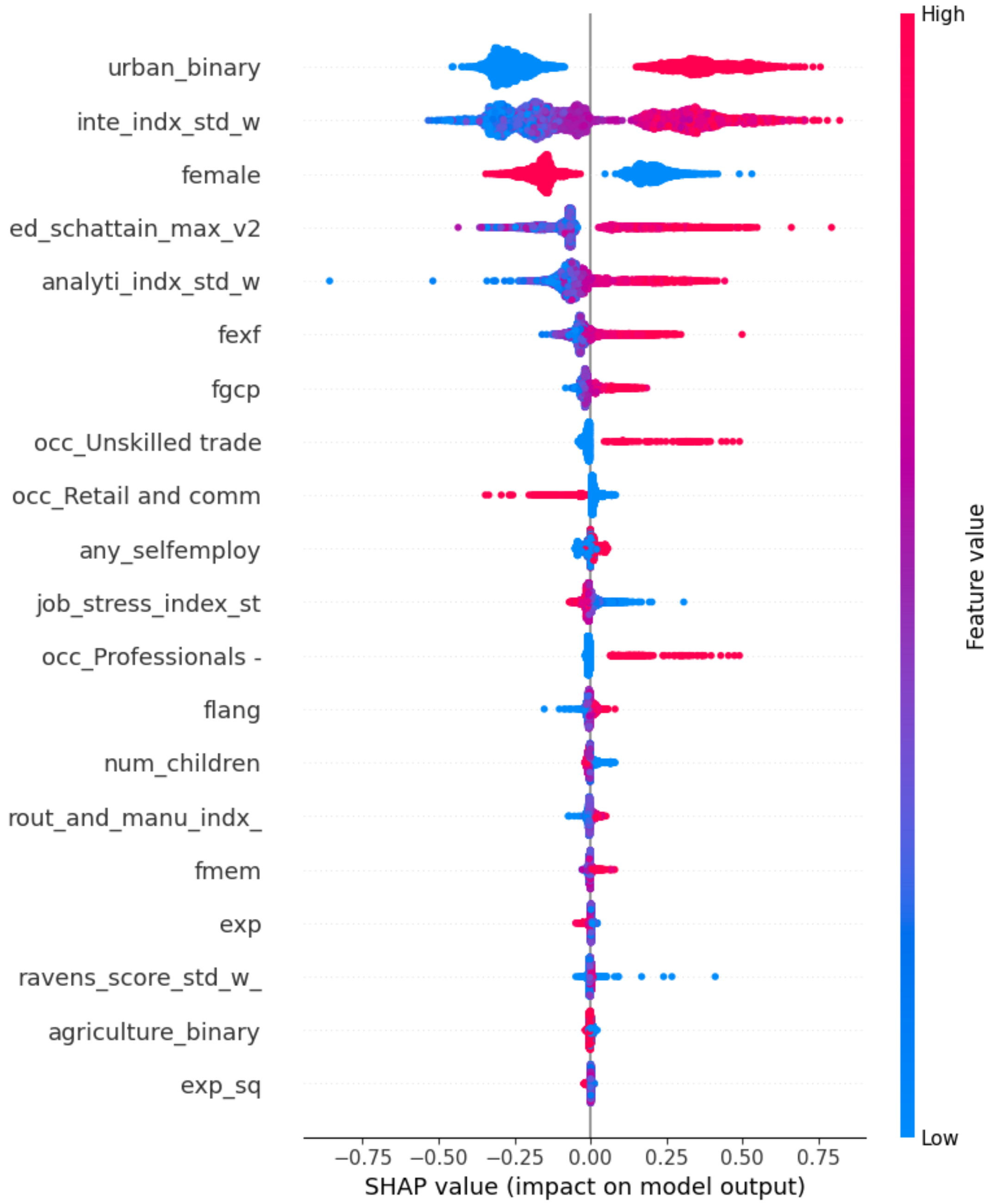


Figure 2.5: SHAP summary plot predicting log wages

Note: Figure shows the impact of key features on predicted log wages, which are derived from a random forest machine learning model. Features are ranked by their average absolute SHAP (SHapley Additive exPlanations) values, with more important predictors appearing higher on the vertical axis. Each dot represents a single observation, and its position on the horizontal axis reflects the SHAP value (i.e., the contribution of the feature to the prediction for that observation). Dot colors indicate feature values, with higher feature values in pink and lower feature values in blue.

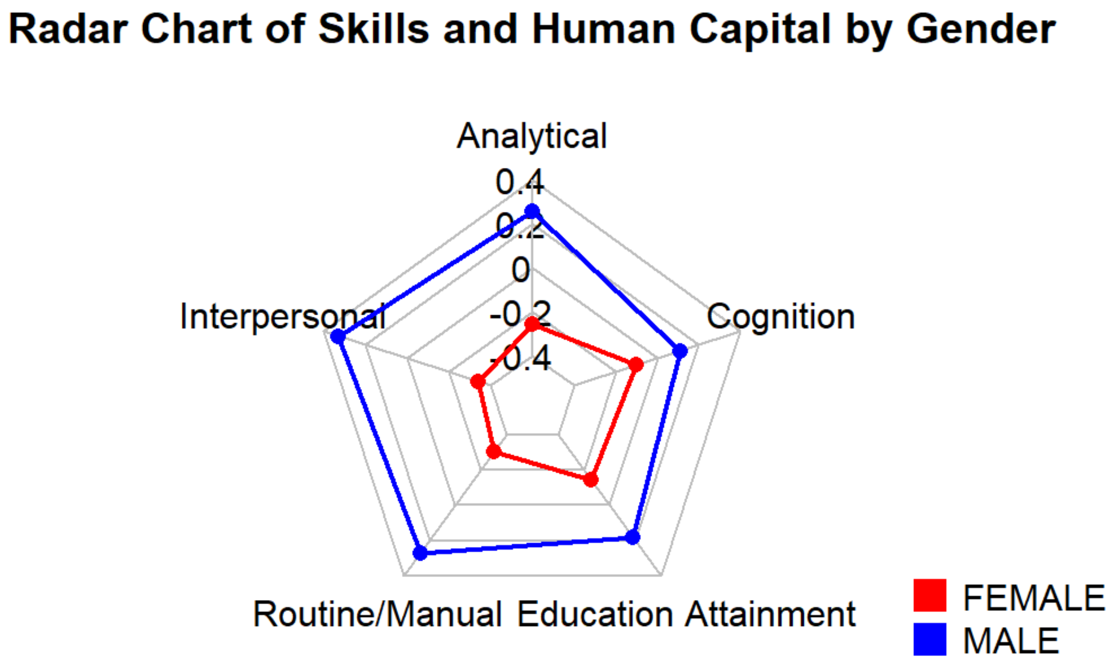


Figure 2.6: Gender Differences in Human Capital Measures

Note: Figure displays (in standard deviations) the average male and female outcomes on educational attainment, general cognition, and the three task indices.

Table 2.1: Summary of labor market outcomes and education by gender

	Mean (Male)	Mean (Female)
Years of Education	9.9	8.9
Hourly Wage (USD)	1.04	0.45
Monthly Earnings (USD)	138.3	49.5
Weekly Hours Worked	46.6	35.3
Any self-employment(%)	46.4	40.5
Any wage employment(%)	52.3	33.2
Not employed past 30 days(%)	5.2	13.2

Note: N=5,877 (excluding 1 observation with missing gender). Outcomes use 1 USD = 140 KSH.

Table 2.2: Top occupations by gender

Occupation	Share Among Female (%)	Share Among Male (%)
Agriculture and Fishing - Farmer	33.2	17.1
Retail and commercial - Hawking/ selling	17.9	5.0
Professionals - Teacher	4.8	5.1
Unskilled trades - Domestic work	8.1	0.8
Agriculture and Fishing - Fishing	0.6	7.7
Skilled & semi-skilled trades - Barber	5.1	0.9
Unskilled trades - Cleaner (other)	4.6	0.9
Unskilled trades - Bicycle/motorbike taxi work	0.0	6.0
Retail and commercial - Own shop	3.0	2.3
Skilled & semi-skilled trades - Tailor	3.9	0.9
Skilled & semi-skilled trades - Mason	0.1	4.9
Unskilled trades - Watchman/ Guard	0.7	3.6
Unskilled trades - Unskilled construction laborer	0.3	4.0
Unskilled trades - Cook/ Chef/ Caterer	2.2	1.6
Other - Other (specify)	1.0	2.9
Skilled & semi-skilled trades - Other skilled construction	0.2	3.8
Unskilled trades - Work in other person's business	1.2	2.4
Professionals - Salaried professional	1.1	2.6
Agriculture and Fishing - Agricultural laborer	1.9	1.6
Professionals - Police/military officer	0.5	2.7

Note: N= 5,878. Table shows the top 20 most common occupations among both men and women in the KLPS-5 sample, in descending order of number of workers. The right two columns show, for each occupation, the share of women and men employed in that occupation respectively.

Table 2.3: Gender wage gap: main regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.78***	-0.62***	-0.58***	-0.50***	-0.48***	-0.41***	-0.41***
	(0.05)	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Educational attainment		0.08***		0.04***	0.02	0.02	0.02
		(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
Experience		-0.13***		-0.03	-0.03	-0.01	-0.02
		(0.05)		(0.04)	(0.04)	(0.04)	(0.04)
Experience^2/100		0.00**		0.00	0.00	0.00	0.00
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Urban		0.62***		0.43***	0.42***	0.37***	0.37***
		(0.06)		(0.06)	(0.06)	(0.06)	(0.06)
Nairobi		0.16**		0.13*	0.14**	0.13**	0.14**
		(0.07)		(0.07)	(0.06)	(0.06)	(0.06)
Agriculture		0.08		-0.05	-0.06	-0.04	-0.03
		(0.06)		(0.06)	(0.06)	(0.06)	(0.05)
Self-employment/business		0.03		0.27***	0.28***	0.24***	0.19***
		(0.05)		(0.05)	(0.05)	(0.05)	(0.05)
General cognition score					0.19***	0.16***	0.04
					(0.03)	(0.03)	(0.27)
# Older adults in HH					0.03	0.03	0.06
					(0.06)	(0.06)	(0.06)
# Children in HH					-0.05***	-0.05***	-0.05***
					(0.01)	(0.01)	(0.01)
Analytical Index						0.10***	0.08***
						(0.03)	(0.03)
Interpersonal index						0.12***	0.10***
						(0.03)	(0.03)
Routine/Manual index						-0.00	0.02
						(0.02)	(0.02)
Job stress index							-0.14***
							(0.02)
N	5063	5063	5061	5061	5061	5061	5061
r2	0.08	0.22	0.31	0.34	0.35	0.36	0.38
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: All columns include the female dummy as an explanatory variable. Column 2 includes standard demographics and education controls. Column 3 includes 49 occupation dummies as controls. Column 4 includes the controls in columns 2 and 3. Column 5 additionally include the general cognition score, number of children, and number of older adults over 65 as controls. Column 6 additionally controls for the three task indices. Column 7 additionally controls for the job stress index, personality measures, and economic preference measures, as well as the five component-specific cognition scores and the lagged Raven's score from KLPS-3. All standard errors are robust.

Table 2.4: Gender gap in hours worked

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-12.10***	-10.59***	-10.07***	-8.73***	-8.58***	-6.98***	-6.83***
	(1.00)	(0.97)	(1.10)	(1.07)	(1.08)	(1.12)	(1.13)
Educational attainment		0.57*		0.40	0.32	0.28	0.23
		(0.31)		(0.29)	(0.29)	(0.29)	(0.30)
Experience		1.25		1.17	1.21	1.48*	1.51*
		(0.99)		(0.89)	(0.89)	(0.89)	(0.87)
Experience^2/100		-0.03		-0.03	-0.03	-0.03	-0.03
		(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
Urban		6.67***		2.47*	2.30*	1.28	1.02
		(1.39)		(1.34)	(1.36)	(1.37)	(1.38)
Nairobi		-1.80		-1.76	-1.69	-1.87	-1.77
		(1.70)		(1.55)	(1.55)	(1.52)	(1.51)
Agriculture		0.12		-3.19**	-3.32***	-2.80**	-2.98**
		(1.28)		(1.25)	(1.26)	(1.27)	(1.26)
Self-employment/business		12.86***		11.65***	11.68***	10.86***	10.80***
		(0.95)		(1.07)	(1.07)	(1.09)	(1.09)
General cognition score					0.63	0.20	15.59**
					(0.55)	(0.56)	(6.27)
# Older adults in HH					-0.79	-0.77	-0.73
					(1.24)	(1.23)	(1.25)
# Children in HH					-0.33	-0.37	-0.37
					(0.26)	(0.26)	(0.26)
Analytical Index						1.36**	1.34**
						(0.61)	(0.61)
Interpersonal index						3.30***	3.32***
						(0.66)	(0.65)
Routine/Manual index						0.37	0.36
						(0.59)	(0.58)
Job stress index							0.10
							(0.54)
N	5773	5773	5772	5772	5772	5772	5772
r2	0.04	0.10	0.21	0.24	0.24	0.25	0.26
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: Please see Table 2.3 for the notes, all of which apply here. The dependent variable is weekly hours worked. Individuals who do not report working any hours are included in estimation and coded as working zero hours. All standard errors are robust.

Table 2.5: Double/Debiased Machine Learning results

	(1)	(2)	(3)	(4)
Dep. Var.: Log Wage				
Female	-0.57***	-0.41***	-0.55***	-0.40***
	(0.04)	(0.04)	(0.04)	(0.04)
Observations	5063	5063	4758	4758
Dep. Var.: Weekly Hours Worked				
Female	-7.07***	-4.43***	-6.78***	-5.07***
	(0.89)	(0.88)	(0.95)	(0.92)
Observations	5773	5773	5120	5120
Controls	Basic	Full	Basic	Full
Class Balance	No	No	Yes	Yes

Note: This table reports results from Double/Debiased Machine Learning (DML). The top row reports the coefficient and standard error on the female indicator in a model where the dependent variable is log wage; in the bottom row the dependent variable is weekly hours worked. Columns 1 and 3 include occupation, education, and demographics as potential confounders; Columns 2 and 4 include the full set of controls. Columns 3 and 4 randomly under-samples women so as to achieve perfect balance between men and women in the analysis sample. All models use a random forest regressor to predict the outcome and a random forest classifier to predict being female. All models use the default 3-fold cross-validation of DML.

Table 2.6: Oaxaca-Blinder decomposition of wages

	Coef.	95% CI Lower	95% CI Upper
With basic controls:			
Difference(Male-Female)	0.780	0.692	0.867
Explained	0.307	0.202	0.411
Unexplained	0.473	0.345	0.601
With full set of controls:			
Difference(Male-Female)	0.780	0.691	0.868
Explained	0.394	0.319	0.468
Unexplained	0.386	0.284	0.488

Note: This table reports the Oaxaxa-Blinder decomposition results of the gender gap in log wages, using male as the reference group. "Basic" controls include eduaction, experience, demographics, and occupation variables. "Full" controls additionally include cognition, personality, economic preferences, job stress, and home care responsibilities.

Table 2.7: Oaxaca-Blinder decomposition of hours worked

	Coef.	95% CI Lower	95% CI Upper
With basic controls:			
Difference(Male-Female)	12.099	10.235	13.964
Explained	1.345	-0.857	3.547
Unexplained	10.754	8.073	13.436
With full set of controls:			
Difference(Male-Female)	12.099	10.201	13.998
Explained	3.635	2.053	5.217
Unexplained	8.464	6.141	10.788

Note: This table reports the Oaxaxa-Blinder decomposition results of the gender gap in hours worked using male as the reference group. "Basic" controls include eduaction, experience, demographics, and occupation variables. "Full" controls additionally include cognition, personality, economic preferences, job stress, and home care responsibilities.

Table 2.8: Gender wage gap: interaction terms

	(1)	(2)	(3)	(4)
Female (intercept)	-0.32	0.43	-0.07	0.93
	(0.21)	(1.02)	(0.19)	(1.10)
Interaction terms with Female				
Urban	-0.07	-0.12	-0.38***	-0.25**
	(0.11)	(0.11)	(0.12)	(0.12)
Nairobi	0.03	0.13	0.20	0.05
	(0.13)	(0.13)	(0.13)	(0.12)
Agriculture	0.08	0.00	-0.24**	-0.13
	(0.11)	(0.10)	(0.11)	(0.11)
Self-employment/business	-0.15*	-0.13	-0.21**	-0.20**
	(0.09)	(0.09)	(0.09)	(0.09)
Educational attainment	-0.02	-0.03	-0.01	-0.02
	(0.02)	(0.03)	(0.01)	(0.03)
General cognition score		0.92*		0.79
		(0.54)		(0.60)
Executive functioning		-0.31		-0.31
		(0.22)		(0.24)
Orientation		-0.01		0.03
		(0.11)		(0.13)
Memory		-0.44		-0.35
		(0.29)		(0.32)
Language		-0.26*		-0.12
		(0.14)		(0.15)
Visuospatial		-0.21*		-0.21*
		(0.12)		(0.12)
Lagged Raven's		0.03		0.00
		(0.05)		(0.05)
Experience		-0.04		-0.08
		(0.08)		(0.09)
Experience^2/100		0.00		0.00
		(0.00)		(0.00)
Analytical Index		-0.14***		-0.15***
		(0.05)		(0.05)
Interpersonal Index		0.18***		0.19***
		(0.05)		(0.06)
Routine/Manual Index		0.04		-0.02
		(0.05)		(0.05)
N	5061	5061	5061	5061
Model	Regression	Regression	DML	DML
Controls	Basic	Full	Basic	Full

Note: Columns 1 and 3 include the basic set of controls a la Column 4 in Table 2.3, along with the listed interaction terms. Columns 3 and 4 include the basic set of controls a la Column 7 in Table 2.3, along with the listed interaction terms. Columns 3 and 4 run DML regressions following the same setup as in Table 2.5. All standard errors are robust.

Table 2.9: Gender gap in employment wage

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.41***	-0.28***	-0.41***	-0.34***	-0.33***	-0.30***	-0.30***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.06)	(0.07)	(0.06)
Educational attainment		0.10***		0.04**	0.02	0.01	0.01
		(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
Experience		-0.16***		-0.09	-0.09	-0.07	-0.08
		(0.06)		(0.05)	(0.05)	(0.05)	(0.06)
Experience^2/100		0.00***		0.00*	0.00*	0.00	0.00*
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Urban		0.50***		0.39***	0.37***	0.34***	0.33***
		(0.08)		(0.08)	(0.07)	(0.07)	(0.07)
Nairobi		0.09		0.12	0.13*	0.12	0.13*
		(0.08)		(0.08)	(0.08)	(0.08)	(0.08)
Agriculture		-0.09		-0.05	-0.06	-0.04	-0.03
		(0.07)		(0.07)	(0.07)	(0.07)	(0.07)
Self-employment/business		0.14*		0.19***	0.19***	0.15**	0.11
		(0.07)		(0.07)	(0.07)	(0.07)	(0.07)
General cognition score					0.15***	0.13***	0.33
					(0.03)	(0.03)	(0.32)
# Older adults in HH					-0.23**	-0.22**	-0.22**
					(0.09)	(0.09)	(0.09)
# Children in HH					-0.04**	-0.04***	-0.04**
					(0.02)	(0.02)	(0.01)
Analytical Index						0.15***	0.12***
						(0.03)	(0.03)
Interpersonal index						0.04	0.03
						(0.04)	(0.04)
Routine/Manual index						-0.08***	-0.05*
						(0.02)	(0.02)
Job stress index							-0.16***
							(0.03)
N	2260	2260	2260	2260	2260	2260	2260
r2	0.04	0.20	0.29	0.32	0.34	0.35	0.39
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: Please see Table 2.3 for the notes, all of which apply here. The dependent variable is wage from employment. Individuals who do not report engaging in any wage employment are excluded from estimation. All standard errors are robust.

Table 2.10: Gender gap in earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-1.11***	-0.90***	-0.87***	-0.76***	-0.72***	-0.61***	-0.60***
	(0.05)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
Educational attainment		0.11***		0.06***	0.03***	0.03**	0.03**
		(0.02)		(0.01)	(0.01)	(0.01)	(0.01)
Experience		-0.11**		-0.00	0.01	0.03	0.02
		(0.05)		(0.04)	(0.04)	(0.04)	(0.04)
Experience^2/100		0.00*		-0.00	-0.00	-0.00	-0.00
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Urban		0.72***		0.43***	0.41***	0.33***	0.32***
		(0.06)		(0.06)	(0.06)	(0.06)	(0.05)
Nairobi		0.11		0.07	0.08	0.07	0.08
		(0.07)		(0.06)	(0.06)	(0.06)	(0.06)
Agriculture		0.29***		0.00	-0.01	0.03	0.04
		(0.06)		(0.05)	(0.05)	(0.05)	(0.05)
Self-employment/business		0.29***		0.51***	0.52***	0.46***	0.42***
		(0.05)		(0.05)	(0.05)	(0.05)	(0.05)
General cognition score					0.21***	0.17***	0.15
					(0.03)	(0.03)	(0.27)
# Older adults in HH					0.02	0.02	0.04
					(0.06)	(0.06)	(0.06)
# Children in HH					-0.06***	-0.06***	-0.06***
					(0.01)	(0.01)	(0.01)
Analytical Index						0.15***	0.13***
						(0.03)	(0.03)
Interpersonal index						0.19***	0.19***
						(0.03)	(0.03)
Routine/Manual index						0.04*	0.06***
						(0.02)	(0.02)
Job stress index							-0.12***
							(0.02)
N	5434	5433	5336	5336	5336	5335	5335
r2	0.12	0.43	0.31	0.38	0.40	0.51	0.52
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: Please see Table 2.3 for the notes, all of which apply here. The dependent variable is log monthly earnings. All standard errors are robust.

Table 2.11: Experiences of discrimination by gender

Panel A: Share experiencing discrimination by type	Mean (Male)	Mean (Female)	Female Coef (No controls)	Female Coef (All controls)
Treated with less courtesy than other people	0.51	0.51	-0.02	-0.03
			(0.02)	(0.02)
Treated with less respect than other people	0.52	0.52	-0.00	-0.02
			(0.02)	(0.02)
Receive poorer service than other people at restaurants/stores	0.31	0.28	-0.04**	-0.06***
			(0.02)	(0.02)
People act as if they think you are not smart	0.46	0.47	0.00	-0.04*
			(0.02)	(0.02)
People act as if they are afraid of you	0.41	0.39	-0.04**	-0.03
			(0.02)	(0.02)
You are threatened or harassed	0.28	0.27	-0.02	-0.07***
			(0.01)	(0.02)
Panel B: Main perceived reason for discrimination	Mean (Male)	Mean (Female)		
Your education or income level	0.48	0.49		
Your Social Group or Family Background	0.18	0.19		
Aspects of your physical appearance	0.12	0.11		
Place of origin	0.06	0.05		
Your gender	< 0.01	0.02		
Your ancestry	0.02	0.01		
Your language ability	0.01	0.01		
Your religion	< 0.01	< 0.01		
Your age	< 0.01	< 0.01		
Your race	< 0.01	< 0.01		
Your sexual orientation	< 0.01	< 0.01		
Your political Orientation	< 0.01	< 0.01		

Note: Panel A shows the male mean and the coefficient on female in regressions where each one of the variables listed on the left-most column is the dependent variable. Each variable is coded as 1 if the respondent reports that the experience have happened to them more than a few times a year or more, and 0 if they report less than once a year or never. The third column includes no covariates; the fourth column includes the full set of covariates used in regression analyses in Table 2.3. Panel B shows the share of individuals, among men and women respectively, who say that the left-most column reason is the main reason why they think discrimination happened to them.

Chapter 3

When Is a Market like a Walmart?

3.1 Introduction

Why are firms in low- and middle-income countries (LMICs) so small, and how much does this matter for consumers? Differences in the firm size distribution between rich and poor countries have long bothered economists. In the retail trading sector specifically, a supplementary hypothesis holds that firm size is limited by the extent of the market. In places where few consumers have private vehicles, demand is spread thin geographically and cannot sustain the types of "superstores" observed in wealthier countries that can capitalize on higher productivity modern retail technologies. These stories imply that alleviating frictions or decreasing transportation costs would increase average firm size and aggregate productivity, and could have substantial impacts on consumer welfare.

We propose that both stories may be true, and yet the gains may be much smaller than expected. We present evidence from Nigerian survey data that small retailers co-located in markets already capture many of the available gains from scale, albeit across rather than within firm boundaries. We then show in a calibrated model that alleviating frictions does lead to increases in average firm size and measured productivity, but the welfare implications are much smaller than they would be in the absence of within-market interactions between firms.

We begin by documenting a set of stylized facts about the wholesale and retail sectors in Lagos, Nigeria using comprehensive data we have collected as part of the fifth round of the Lagos Trader Survey (LTS) covering 1,505 micro and small wholesale and retail businesses in major markets in Lagos, Nigeria, a city with a population of over 20 million and the gateway to the rest of Nigeria. These traders source manufactured consumer goods such as apparel, electronics, and home appliances from either abroad or wholesalers within Nigeria from all over the world, and sell to either end consumers or other retailers.

We document three key patterns in our data. First, firms are indeed small, but are co-located with many other small firms selling the same products, in markets that serve a much larger amount of spatially aggregated demand. Second, firms frequently source and sell goods within their own market. 75% of firms sell to other trading businesses within the same market, and 47% source from others within the market. Such within-market aggregation can generate meaningful volumes of trade for wholesalers: the median firm in the LTS sample earns \$7,000 annual revenue and is surrounded by 15 other shops selling the same or similar products in their own market. Third, larger firms with more workers face increasing labor costs suggetive of span of control problems. Owners spend more time on management, but they do not employ sophisticated management practices. Most often, owners supervise workers by simply watching them; few solicit feedback from customers or other workers. Owners at larger firms tend to spend more time on management, pay a higher hourly wage to workers, and earn lower profit per worker.

In order to measure the welfare consequences of small retail firms, we need to take a stand on why these firms are small in the first place. In our model, firms benefit from internal economies of scale in sourcing and operation, but they also face management costs to prevent employees from shirking or stealing. As such, firms need to pay efficiency wages and dedicate owner's time to monitoring so as to incentivize effort from workers. Management costs increase with respect to firm size as owner's time becomes more scarce. Furthermore, firms can choose whether they would like to engage in retail, wholesale, or both. Wholesale firms are able to generate more sales per employee and thus face lower costs of monitoring and enjoy larger economies of scale at the expense of lower markups.

We formalize this intuition in a quantifiable model. We microfound management costs through a model of efficiency wages where the owner has to spend time on monitoring their workers. More time spent on monitoring reduces the efficiency wages that must be paid, but comes at the expense of scarce owner time, as every owner is endowed with a fixed amount of time, and time spent managing comes at the cost of time spent sourcing or selling. Our model features a continuum of firms which vary in their productivity. Firms may choose to enter wholesale, paying higher fixed costs to buy directly from the manufacturer, choose a quantity, which determines both their variable costs and their management costs, and then resell on wholesale markets for an endogenous wholesale price. Alternatively, firms may choose to enter retail and pick a quantity to source from wholesalers paying endogenous wholesale prices and the corresponding management cost. In equilibrium, higher-productivity firms engage in wholesale, lower-productivity firms engage in retail, and the least productive entrants exit because they earn negative profits. We derive the first-order conditions for the three kinds of firms, as well as the wholesale and retail price indices. In future work, once we finalize the model and calibrate the parameters, we can conduct counterfactual exercises that examine the effects of reducing span of control problems (e.g., via introduction of more efficient monitoring technology), permitting us to endogenously obtain larger retailers ("Walmarts") and assess how this affects consumer prices and the overall gains from retailer consolidation.

We contribute to the literature on the firm size distribution in developing countries (Hsieh and Klenow, 2014; Hsieh and Olken, 2014). Firms in low- and middle- income countries are often small and spatially close to each other within an industry Banerjee et al. (2022), Bassi, Muoio, et al. (2022), and Vitali (2024). Existing literature has documented reasons for the lack of firm expansion and consolidation, such as lack of capital (Mel, McKenzie, and

Woodruff, 2008; Mel, McKenzie, and Woodruff, 2009; Mel, McKenzie, and Woodruff, 2011; Fafchamps et al., 2014), constraints in finding suppliers and customers (Jensen and Miller, 2018; Aker, Blumenstock, and Dillon, 2020), and information constraints in labor reallocation (Hardy et al., 2024). In this project, we show that the presence of chains of intermediation and span of control problems lead to relatively modest gains from firm consolidation, which can explain the firm size distribution that we observe empirically. Related to our work, Bassi, Lee, et al. (2023) study the manufacturing sector in Uganda and find that the largest firms in their setting more closely resemble a collection of self-employed individuals due to lack of labor specialization; as a result, the gains from firm consolidation are limited. We study the retail sector and provide alternative but complementary explanations that contribute to limited gains from firm consolidation.

We also extend prior work in Startz (2024) and Grant and Startz (2024) and join a small set of papers focused on how wholesale and retail distribution influence consumer welfare in developing countries (Atkin and Donaldson, 2015; Lagakos, 2016). In middle-income countries such as Mexico, retail globalization has led to the entry of foreign big-box chains that were able to offer lower consumer prices and increase competition in the retail sector (Atkin, Faber, and Gonzalez-Navarro, 2018). However, such changes have yet to take off to a significant extent in low-income countries like Nigeria. We extend this literature by examining how unique features of the retail sector in low-income countries—the co-location of small firms and the presence of chains of intermediation together with span of control problems—contribute to the observed equilibrium by allowing firms to take advantage of economies of scale. By quantifying these dynamics, we provide a deeper understanding of how the structure of retail markets impacts consumer welfare in low-income settings.

3.2 Data and Summary Statistics

The Lagos Trader Survey (LTS) is a panel survey of wholesale and retail traders in major markets in Lagos, Nigeria. The fifth round of LTS (hereafter, LTS5) was conducted between July 2023 and August 2024, and includes an original sample of 559 respondents surveyed in previous rounds, as well as a new sample of 946 respondents, totaling 1,505. Respondents from both the original and new LTS sample were identified through a census of over 68,000 shops in commercial areas of Lagos, and interviews were conducted with traders whose shops were randomly sampled from the census.[1] For both the original and new sample, we implemented a random sampling approach, where enumerators first conducted a complete listing of all establishments in a market, and then randomly sampled shops based on specific instructions given (e.g., "every third shop after you turn right from the entrance on the ground floor"). The survey includes detailed questions on firm and owner characteristics, sourcing and selling patterns, hired labor, owner and workers' time use patterns, interaction

[1] The original census was conducted between 2014 and 2015. The 559 respondents from the original census already responded to previous rounds of the panel survey and were re-visited for LTS5. We conducted an updated census in 2023 to sample new respondents.

among traders in the same market, management practices, capital stock, cost breakdown, revenue, and profit. To reduce the length of the survey, some sections of the survey were randomly administered to one-third of the respondents.

To collect transaction-level information, we ask respondents in detail about the trips they took or orders they placed (without physically traveling) for sourcing their goods, either inside or outside Nigeria, including sourcing location, quantities, transportation costs, costs paid for the goods, and any other costs involved. The survey also includes a unique set of variables that describe how transactions are conducted, such as whether the buyer traveled to the source country and details about their trip if they did, method of payment, fraction of payment made pre-shipment, mode of communication with suppliers, use of hired agents, and whether there were any delays or defects compared to expected product delivery. For more information about the data, please see Startz (2024).

We ask firm owners a comprehensive set of questions about their own market. Specifically, we ask respondents whether they source from or sell to any other traders within a 10-minute walk from them, and if so, the percentage sourced from or sold to those traders. To understand aggregate market size, we also ask respondents to estimate how many other shops within a 10-minute walk sell products that are the same as or similar to theirs. Furthermore, we ask traders if they share any physical shop space or equipment (e.g., generators) with other traders, if they collaborate with other traders (e.g., sourcing goods together on the same order or giving and receiving customer referrals), and if they engage in discussions with other traders about business topics.

To understand how owners and workers spend time on different tasks, we ask respondents, who are business owners, how many hours per week they spend on their business, and provide a percentage breakdown of their time into four categories: attending to customers; sourcing goods, which includes looking for new suppliers or products, making purchases, and taking deliveries; management, which includes record keeping, loans, hiring or supervising workers; and other tasks, such as examining inventory and cleaning the shop. We also ask respondents to estimate how many hours their workers spend on their business per week, and to provide a percentage breakdown across the same four time-use categories. We were able to capture this information for the vast majority of the respondents: 98% of them provided valid responses to the time-use questions.

The sample includes any trader dealing in manufactured consumer goods, excluding food, and excludes street vendors. The firms in our sample sell a diverse range of products. Figure 3.1 shows a bar chart of product types and the share of firms selling each product. The most prominent categories include clothing, shoe/bag/leather, electronic accessories, jewelry, electrical hardware, and baby/children goods (e.g., diapers). Firms tend to concentrate on one product category: 96% of firms sell products within only one category. In our sample, 13% of the firms do only wholesale, 13% do only retail, and 74% do both. Among firms that do both, on average, 45% of their sales come from retail. The engagement in both wholesale and retail businesses is common among traders in Lagos markets: Most function both as wholesalers to downstream firms and retailers to individual customers who visit their shops.

Table 3.2 summarizes key statistics for the entire sample as well as its three segments:

retail-only firms, wholesale-only firms, and firms that conduct both wholesale and retail businesses. Overall, 51% of the sample are importers; however, the share is much lower among retail-only firms, at 16%. The firms in our sample are small in terms of employment: 52% have zero employees and 88% have fewer than 3 employees. Figure 3.2 displays the histogram of the number of employees per shop. Firms that do wholesale tend to have more employees than retail-only firms. However, even though most respondents are owner-operators or only have a few employees, their businesses handle a large volume of goods, which is perhaps unsurprising for established businesses in the wholesale and retail sector. The average annual revenue per firm is $27,478, and the median revenue is $7,000.[2] The average annual take-home profit per firm is $2797, and the median profit is $1,000. The average firm owns $42,580 worth of business property (which includes goods in the inventory, shop space, warehouses, equipment, etc.), and the number for the median firm is $13,750. These results show that in terms of revenue and profit, the firms in LTS are larger than the typical microentrepreneur-owned firm in a low-income country.

3.3 Descriptive Facts

In this section we document several stylized facts from the Lagos Trader Survey. We show that firms frequently source within markets, allowing them to take advantage of the economies of scale offered by upstream wholesalers. Firms within a market also collaborate in ways that resemble departments within a large firm. We also document evidence showing that larger firms face a larger burden of span-of-control problems, since they need to supervise more employees but the owner's time is limited.

Fact 1: Many firms sell and source within the same market. The last two rows of Table 3.2 show the share of firms that sell to or source from other shops within the market, respectively.[3] Overall, 47% of firms source from other shops within the market. Unsurprisingly, the share is smaller among firms that only do wholesale (39%) than among those that only do retail (49%) or both (48%). Similarly, 75% of firms sell to other traders within the market, though the percentage varies: the number is smaller (48%) among retail-only firms than wholesale-only firms (83%) and firms that do both wholesale and retail (79%). For firms that source within the market, the percentage of goods sourced from those suppliers is significant: on average, 41% of their total value of goods are sourced from within-market suppliers (the median is 25%). Similarly, for firms that sell within the market, on average, 22% of their total sales are to other traders within the market (the median is 20%). When asked why they source from shops within the same market, the most frequently mentioned reason is that it reduces the stress and transportation costs of sourcing from places farther away. Within-market sourcing applies to both domestic and imported goods. In many cases,

[2]Unless noted otherwise, all numbers in USD in the paper use a 1:1000 exchange rate, which was similar to the market exchange rate in January and February 2024.

[3]Specifically, we define "within the market" as any shops within a 10-minute from the respondent's business.

firms buying imported goods have established relationships with importers located in the same market and source directly from them. Another type of retailer within markets is mobile vendors (i.e., hawkers). However, firms in the LTS5 sample do not sell extensively to mobile vendors. Only 17% of firms sell to hawkers, and among those that do, an average of 14% of their sales (the median is 10%) goes to them.

As shown in Grant and Startz (2024), this aggregation of demand through wholesalers/ importers allows smaller firms to save on fixed cost and obtain access to a larger variety of products from faraway places, while paying slightly higher variable costs because the wholesalers/importers charge a markup. Indeed, the demand aggregated from multiple retailers can be quite substantial. To estimate the possible degree of demand aggregation, we ask business owners to estimate the number of shops selling the same or similar products within their own market. Figure 3.3 displays the distribution of participants' estimates for the number of shops selling the same or similar products within their market. The average estimate is 32 and the median is 15, which aligns with our field observation that multiple shops selling the same or similar products tend to cluster together. In some areas, aggregation produces clusters of substantial size. For example, in Computer Village, an area of about 0.4 square kilometers in size, there are approximately 3,000 shops selling electronics and electronic accessories.[4]

Fact 2: Firms in the same market collaborate with each other on sourcing, referral, and equipment sharing. 97% of traders report that they interact socially with other businesses in the same market. Table 3.3 shows the percentage of traders who engage in various kinds of collaborative activities with other businesses in the same market. 58% of firms either give or receive customer referrals; anecdotally, referrals often take place when a business does not have the product that the customer asked for so they will refer the customer to nearby shops. Business owners also collaborate with each other on several different tasks: 20% give or receive introduction to a supplier or agent, 18% borrow or lend stocks, 9% source products on the same order with others, 7% share shop space or stores. Additionally, 21% of respondents report giving or receiving a gift or loan from other traders. 30% of respondents share equipment with other shops. Among those, 89% share generators, 13% share fans, and 5% share calculators. The number of respondents sharing other kinds of equipment is negligible. These observations show that firms in a market collaborate in ways that benefit each other, sometimes resembling different departments within a large retail company.

Fact 3: Larger firms incur larger costs from span of control problems. We document the time use patterns of owners and workers across the following categories of activities: attending to customers, sourcing goods, management (of both workers and businesses, such as record keeping and dealing with loans), and other tasks, which include cleaning, taking stock of inventories, etc. Table 3.4 shows the distribution of time use, and Figure 3.4 summarizes the means, for the following three groups of individuals: firm owners with zero employees, firm owners with one or more employees, and workers. On average, owners of firms with zero employees spend 61 percent of their time attending to customers, 20 percent

[4] Estimate based on observations from the field team.

sourcing goods, and 9 percent doing management. Workers spend a similar share of time (62 percent) on attending to customers, though they tend to spend less time sourcing goods (10 percent) and management (7 percent). This difference occurs because workers spend more time taking stock and cleaning, which is expected. Turning to owners of firms with one or more employees, we see that compared to solo entrepreneurs, they spend a lower share of their time on attending to customers (51 percent) and a higher share on management (13 percent) and sourcing (25 percent). These differences are statistically significant at the 1% level, and they potentially reflect the differences in the roles owners take as they transition from solo entrepreneurs to owner-managers. Owners of firms with one or more employees likely delegate customer-facing tasks to employees, so that they can focus more on sourcing goods and managing workers.

Although owners at larger firms tend to spend a higher share of their time on management, their methods for supervising workers are not very sophisticated. When asked about the approaches they use to supervise their workers, 80% of owners with any employees say they supervise workers by watching them. Only 12% say they ask their manager or other workers to watch their workers, 10% use a camera, and 8% say they ask customers or neighbors for feedback on their workers. The lack of sophisticated supervision techniques likely introduces span of control problems because owners likely are not capable of effectively supervising workers by just watching them, as owners are often away from their shops and engage in a variety of other tasks, such as calling suppliers and clients, even when they are physically present.

We also estimate the relationship between firm size and measures of revenue, costs, and profit to understand the costs of scaling up for firms with at least a non-zero number of employees. Table 3.5 shows the regression results of those measures on the number of employees, among firms with one or more employees. All regressions control for an indicator of whether the firm imports, an indicator for the survey round, indicators of wholesale only/retail only/both, enumerator FEs and product category FEs.[5] We find that larger firms tend to pay more per hour for their workers. The magnitude is meaningful: For example, adding an extra employee is associated with a \$1.40 increase in the hourly wage that the owner needs to pay, which is substantial in light of the median hourly wage in this sample being \$6.63. Owners at larger firms also spend a higher share of their time on management, and are more likely to use sophisticated supervision methods (defined as anything other than "watching my workers"). Revenue and profit per worker are also negatively associated with firm size—adding an extra worker is associated with a 15 log point decrease in revenue and 13 log point decrease in profit. While the results are from a cross section and could be affected by unobserved omitted variables, we believe these estimates provide suggestive evidence that larger firms incur higher costs as they grow. As firms expand, they need to pay a higher hourly wage, spend more time supervising their workers, and earn lower profits per worker.

[5]Original LTS respondents were surveyed in Round 1 between July and November 2023; new respondents were surveyed in Round 2 between April and August 2024.

Distribution of firm productivity. We conduct a preliminary exercise to estimate the distribution of firm productivity in the LTS sample. We specify the following production function:

$$\ln Y_i = \beta_0 + \beta_1 \ln K_i + \beta_2 \ln L_i + \beta_3 \ln Input_i + 1(Import)_i + WholesaleRetail_i + \alpha_j + \alpha_k + \epsilon_i$$

Here Y_i is log total revenue, K_i is log capital (i.e., total value of their business property), L_i is the log of the number of workers (including the owner), $Input_i$ is log total input amount, which is total cost minus cost for goods purchased and equipment bought.[6] All variables are measured for the last calendar year (2023). $WholesaleRetail_i$ is an indicator for whether the firm does only wholesale, only retail, or both. α_j and α_k are enumerator and product type fixed effects, i.e., the 14 types of products that the businesses sell. The estimation results are displayed in Table 3.6. We also include an indicator for the survey round. The regression in Column 1 does not include any controls; Column 2 includes the full set of controls as specified above.

We obtain the residuals (log TFP) from the regression with controls and plot its distribution. Figure 3.5(a) shows the kernel density plot. The distribution of residuals closely resembles a normal distribution. Figure 3.5(b) shows the distribution of TFP (i.e., log(residual)), which seems to follow a log-normal distribution.

3.4 Model

In this section, we present an outline of a model that characterizes the retail sector in Lagos, Nigeria. This is a model of retail with three key features:

1. Internal economies of scale, as captured by fixed costs of entering wholesale and retail;
2. Span of control problems, as captured by upward sloping cost curves with respect to firm size (i.e., number of workers), which create diseconomies of scale
3. The existence of wholesalers, with the potential for one or two intermediary chains.

These features are embedded in a chain structure similar to the one in Grant and Startz (2024) characterized by CES demands at every level. This framework permits genuine differences in firm productivity plus a wedge—namely, span of control problems—which leads to inefficiency. Relaxing this wedge will lead to a reallocation towards better firms and greater aggregate welfare. Furthermore, as this wedge relaxes, it will tend to drive pure retailers out of the market through the standard reallocation force, by pushing up the cutoff productivity to enter retailing and pushing down the cutoff for firms to enter both wholesale and retail. This leads to a decline in wholesale demand that will tend to push businesses out

[6] We subtract these two categories from total cost because they might already be counted in the total value of the business. We run a specification with those two cost categories included and the results are very similar.

of wholesaling and into retail, even as the most productive firms still tend to sell relatively more through wholesale than retail.

Retail and Wholesale Demand

Setup

There is a continuum of atomistic firms size M. Firms have productivity φ distributed Beta with CDF defined over the interval $[0,1]$, so that firms never sell below the price they buy at.

$$G(\varphi) = \varphi^{a-1}(1-\varphi)^{b-1}\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

where $\Gamma(\cdot)$ is the gamma function, where a, b are shape parameters. Note this distribution has mean $\frac{a}{a+b}$ and variance $\frac{ab}{(a+b)^2(a+b+1)}$.

Consumer utility is given by

$$U = X_0 + E_r \ln X_1 \tag{3.1}$$

$$X_1 = \left(\int_{\omega\in\Omega_r} q(\omega)^{\frac{\sigma_r-1}{\sigma_r}} d\omega\right)^{\frac{\sigma_r}{\sigma_r-1}} \tag{3.2}$$

where E_r is expenditure on the good sold retail denoted X_1, X_0 is some outside good with price 1, and Ω_r is the set of retail sellers which is an endogenous object. This yields the usual constant elasticity of demand given by σ_r; to simplify notations, we denote the quantity sold by a given firm by $q = A_r p^{-\sigma_r}$ where $A_r \equiv \frac{E}{P_r^{1-\sigma_r}}$ and P_r is the retail price index.[7]

Firms pay fixed costs F_r and F_w to enter retail and wholesale, respectively. We define $Q \equiv c_w q_w + c_r q_r$ as an ingredient in firms' cost functions. We assume $c_r \geq c_w \geq 1$, $\omega \geq 1$, and the fixed costs sufficiently large such that all firms sell at least 1 unit, i.e. $q_w + q_r \geq 1$. Together, these assumptions deliver that no firm has marginal costs below the purchase price in equilibrium. Note that either q_r or q_w may (but need not be) be zero. There are three cases:

1. If a firm enters wholesale, it buys from the manufacturer at cost p_m. Its cost function is given by $C(q_w, \varphi) = F_w + p_m\left(q_w + \frac{Q^\omega}{\varphi}\right)$.

2. If a firm enters wholesale and retail, it buys from the manufacturer at cost p_m. Its cost function is given by $C(q_w, q_r, \varphi) = F_r + F_w + p_m\left(q_r + q_w + \frac{Q^\omega}{\varphi}\right)$.

3. If a firm enters just retail, it has a discrete choice across the set of available wholesalers. If it chooses wholesaler κ with price $p_w(\kappa)$, it gets a match-specific draw with wholesaler κ, $\epsilon(\kappa)$, distributed Fréchet with minimum $m = 0$, shape parameter α, and scale

[7] We can readily extend this to multiple retail sectors.

parameter $s = \left(\Omega_w^{\frac{1}{\alpha}} \Gamma\left(1-\frac{1}{\alpha}\right)\right)^{-1}$ where Ω_w is the measure of wholesale sellers.[8] The firm then has a cost function given by given by $C\left(q_r, \varphi, \kappa\right) = F_r + p_w\left(\kappa\right)\left(q_r + \frac{Q^\omega}{\varphi}\right)$, where the shock $\epsilon\left(\kappa\right)$ is a multiplicative shock on variable profits.

Firm decisions

Retail only pricing and profits. Retailer with productivity φ and buying from wholesaler κ faces the profit maximization problem

$$\pi_r\left(q_r, \varphi, \kappa\right) = \epsilon\left(\kappa\right)\left[p_r\left(\varphi, \kappa, q_r\right) q_r - p_w\left(\kappa\right)\left(q_r + \frac{c_r}{\varphi} Q^\omega\right)\right] - F_r$$

yielding a first-order condition for optimal quantity (and the resulting optimal choice of quantity and price, where we denote that the firm does retail only using ˆ)

$$\begin{aligned}
\frac{\partial \pi_r}{\partial q_r} = 0 &= \epsilon\left(\kappa\right)\left[\frac{\sigma_r - 1}{\sigma_r}\left(\frac{q_r}{A_r}\right)^{-\frac{1}{\sigma_r}} - p_w\left(\kappa\right)\left(1 + \omega\frac{Q^\omega}{\varphi q_r}\right)\right] \\
&= \frac{\sigma_r - 1}{\sigma_r}\left(\frac{q_r}{A_r}\right)^{-\frac{1}{\sigma_r}} - p_w\left(\kappa\right)\left(1 + \omega\frac{Q^\omega}{\varphi q_r}\right) \\
0 &= \frac{\sigma_r - 1}{\sigma_r} p_r\left(\varphi, \kappa, q_r\right) q_r - p_w\left(\kappa\right)\left(q_r + \omega\frac{Q^\omega}{\varphi}\right) \\
\frac{1}{\sigma_r} p_r\left(\varphi, \kappa, q_r\right) q_r + \left(\omega - 1\right) p_w\left(\kappa\right)\frac{Q^\omega}{\varphi} &= p_r\left(\varphi, \kappa, q_r\right) q_r - p_w\left(\kappa\right)\left(q_r + \frac{Q^\omega}{\varphi}\right)
\end{aligned}$$

$$\begin{aligned}
q_r &= A_r^{\frac{1}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r - 1}{\sigma_r \omega}\frac{\varphi_\rho}{p_w\left(\kappa\right) c_r^\omega}\right)^{\frac{\sigma_r}{\sigma_r(\omega-1)+1}} \\
\hat{p}_r\left(\rho, \kappa\right) &= A_r^{\frac{\omega-1}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r \omega}{\sigma_r - 1}\frac{p_w\left(\kappa\right) c_r^\omega}{\varphi_\rho}\right)^{\frac{1}{\sigma_r(\omega-1)+1}}
\end{aligned}$$

This implies that if the firm selects supplier κ it will have profits given by

$$\begin{aligned}
\pi_r\left(\rho, \kappa\right) =& \epsilon\left(\kappa\right) p_w\left(\kappa\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}} A_r^{\frac{\omega}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r - 1}{\sigma_r}\frac{\varphi_\rho}{\omega}\right)^{\frac{\sigma_r - 1}{\sigma_r(\omega-1)+1}} \\
&\left[c_r^{\frac{1-\omega\sigma_r}{\sigma_r(\omega-1)+1}} - c_r^{\omega(1-\omega)}\left(\frac{\sigma_r - 1}{\sigma_r}\frac{1}{\omega}\right)\right] - F_r
\end{aligned}$$

Suppose there are Ω_w measure of wholesalers. Then since $\epsilon\left(\omega\right)$ is Fréchet with shape α and scale $\left(\Omega_w^{\frac{1}{\alpha}} \Gamma\left(1-\frac{1}{\alpha}\right)\right)^{-1}$, the variable profits will be distributed with shape α and scale

[8]This scale parameter was chosen to avoid giving incentive for firms to source wholesale apart from differences in fixed and marginal costs.

$Bp\left(\kappa\right)^{1-\sigma_r}\left(\Omega_w^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right)\right)^{-1}$ where

$$\begin{aligned}B\left(\rho\right)\equiv &A_r^{\frac{\omega}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r-1}{\sigma_r}\frac{\varphi_\rho}{\omega}\right)^{\frac{\sigma_r-1}{\sigma_r(\omega-1)+1}}\\ &\left[c_r^{\frac{1-\omega\sigma_r}{\sigma_r(\omega-1)+1}}-c_r^{\omega(1-\omega)}\left(\frac{\sigma_r-1}{\sigma_r\omega}\right)\right]\end{aligned}$$

In turn, the maximum across all potential suppliers will be Fréchet with shape α, minimum of 0, and scale of
$B\left(\rho\right)\left(\Omega_w^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right)\right)^{-1}\left(\int_{\omega\in\Omega_w}p\left(\kappa\right)^{\alpha(1-\sigma_r)}d\kappa\right)^{\frac{1}{\alpha}}$. This implies an expected payout of

$$\begin{aligned}\mathbb{E}\left[\max_\kappa\pi_r\left(\rho,\kappa\right)\right]&=B\left(\rho\right)\left(\Omega_w^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right)\right)^{-1}\left(\int\limits_{\omega\in\Omega_w}p\left(\kappa\right)^{\alpha(1-\sigma_r)}d\omega\right)^{\frac{1}{\alpha}}\cdot\Gamma\left(1-\frac{1}{\alpha}\right)\\&=B\left(\rho\right)\Omega_w^{-\frac{1}{\alpha}}\left(\int\limits_{\omega\in\Omega_w}p\left(\kappa\right)^{\alpha(1-\sigma_r)}d\omega\right)^{\frac{1}{\alpha}}\end{aligned}$$

Retailer only choice of wholesaler and wholesale price indices. Firms that only do retail must choose a wholesaler. They will do so in a way that maximizes profits:

$$\begin{aligned}\arg\max_\kappa\pi_r\left(\rho,\kappa\right)&=\arg\max_\kappa\left[\epsilon\left(\kappa\right)p_w\left(\kappa\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}B-F_r\right]\\&=\arg\max_\kappa\left[\epsilon\left(\kappa\right)p_w\left(\kappa\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}B\right]\\&=\arg\max_\kappa\left[\ln\epsilon\left(\kappa\right)+\left(\frac{1-\sigma_r}{\sigma_r\left(\omega-1\right)+1}\right)\ln p\left(\kappa\right)+\ln B\right]\\&=\arg\max_\kappa\left[\frac{\sigma_r\left(\omega-1\right)+1}{\sigma_r-1}\ln\epsilon\left(\kappa\right)-\ln p\left(\kappa\right)\right]\end{aligned}$$

Since $\epsilon\left(\kappa\right)$ is distributed Fréchet, $\ln\epsilon\left(\kappa\right)$ is distributed Gumbel with location parameter $-\ln\left(\Omega_w^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right)\right)$ and scale parameter α^{-1}, and thus $\frac{\sigma_r(\omega-1)+1}{\sigma_r-1}\ln\epsilon\left(\kappa\right)$ is distributed Gumbel with location $-\frac{\sigma_r(\omega-1)+1}{\sigma_r-1}\ln\left(\Omega_w^{\frac{1}{\alpha}}\Gamma\left(1-\frac{1}{\alpha}\right)\right)$ and scale $\left(\alpha\left(\frac{\sigma_r-1}{\sigma_r(\omega-1)+1}\right)\right)^{-1}$.

Following Anderson, De Palma, and Thisse (1987), this implies wholesale firms face CES demands with wholesale elasticity σ_w given by

$$\sigma_w=\alpha\left(\frac{\sigma_r-1}{\sigma_r\left(\omega-1\right)+1}\right)+1$$

Furthermore, the expected retail price from this type of firm (to the power of $1-\sigma_r$ to capture its contribution to the price index) will be

$$\begin{aligned}
\mathbb{E}\left[\hat{p}_r(\rho,\kappa)^{1-\sigma_r}\right] &= A_r^{\frac{(\omega-1)(1-\sigma_r)}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r\omega}{\sigma_r-1}\frac{c_r^{\omega}}{\varphi_\rho}\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}\mathbb{E}\left[p_w(\kappa)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}\right] \\
&= A_r^{\frac{(\omega-1)(1-\sigma_r)}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r\omega}{\sigma_r-1}\frac{c_r^{\omega}}{\varphi_\rho}\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}\mathbb{E}\left[p_w(\kappa)^{\frac{1-\sigma_w}{\alpha}}\right] \\
&= A_r^{\frac{(\omega-1)(1-\sigma_r)}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r\omega}{\sigma_r-1}\frac{c_r^{\omega}}{\varphi_\rho}\right)^{\frac{1-\sigma_r}{\sigma_r(\omega-1)+1}}\mathbb{E}\left[p_w(\kappa)^{\frac{1-\sigma_w}{\alpha}}\right]
\end{aligned}$$

Therefore, if we define the wholesale pricing index,[9] P_w, by

$$P_w^{1-\sigma_w} \equiv \frac{1}{\Omega_w}\int\limits_{\varphi\in\Omega_w} p_w(\varphi)^{\frac{1-\sigma_w}{\alpha}}\,dG(\varphi)$$

then we can express the expected retail price as

$$\mathbb{E}\left[\hat{p}_r(\rho,\kappa)^{1-\sigma_r}\right] = A_r^{\frac{\omega-1}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_r\omega}{\sigma_r-1}\frac{c_r^{\omega}}{\varphi_\rho}\right)^{\frac{1}{\sigma_r(\omega-1)+1}} P_w^{1-\sigma_w}$$

And if we define the wholesale profit index, $\tilde{P}_w$,[10] by

$$\tilde{P}_w^{\sigma_w-1} \equiv \Omega_w^{-\frac{1}{\alpha}}\left(\int\limits_{\varphi\in\Omega_w} p(\varphi)^{\alpha(1-\sigma_r)}\,dG(\varphi)\right)^{\frac{1}{\alpha}}$$

then we can express the expected profits of a retailer by

$$\mathbb{E}\left[\max_{\kappa}\pi_r(\rho,\kappa)\right] = B(\rho)\,\tilde{P}_w^{\sigma_w-1}$$

Wholesaler only pricing. Wholesale only firms have profit function (where we use a ˆ on price to denote that these firms are wholesale only)

$$\pi_w(q_w,\rho) = \hat{p}_w(\rho,q_w)\,q_w - p_m\frac{Q^{\omega}}{\varphi_\rho} - F_r$$

and as established in Section 2.2, they face CES demands with elasticity σ_w. We use A_w to capture the demand shifter faced by wholesalers—this object is exogenous from the perspective of a (small) wholesaler but depends on the full set of sourcing decisions, etc.

[9] The wholesale pricing index is the contribution of wholesale prices to retail prices and is distinct from the wholesale profit index as defined below.

[10] The wholesale profit index is the contribution of wholesale prices to retail prices and is distinct from the wholesale profit index as defined above

Taken together, this yields a FOC for optimal quantity and the resulting optimal choice of quantity and price

$$
\begin{aligned}
\frac{\partial \pi_w}{\partial q_w} = 0 &= \frac{\sigma_w - 1}{\sigma_w}\left(\frac{q_w}{A_w}\right)^{-\frac{1}{\sigma_w}} - \omega c_w p_m \frac{Q^{\omega-1}}{\varphi_\rho} \\
&= \frac{\sigma_w - 1}{\sigma_w}\left(\frac{q_w}{A_w}\right)^{-\frac{1}{\sigma_w}} - \omega p_m \frac{c_w^{\omega} q_w^{\omega-1}}{\varphi_\rho} \\
q_w &= A_w^{\frac{1}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_w - 1}{\sigma_w \omega}\frac{\varphi_\rho}{p_m c_w^{\omega}}\right)^{\frac{\sigma_w}{\sigma_w(\omega-1)+1}} \\
\hat{p}_w(\rho) &= A_w^{\frac{\omega-1}{\sigma_r(\omega-1)+1}}\left(\frac{\sigma_w \omega}{\sigma_w - 1}\frac{p_m c_w^{\omega}}{\varphi_\rho}\right)^{\frac{1}{\sigma_w(\omega-1)+1}}
\end{aligned}
$$

Wholesaler and retailer pricing. Firms which engage in both wholesale and retail have profit function

$$
\pi_{wr}(q_w, q_r, \rho) = p_w(\rho) q_w + p_r(\rho) q_r - p_m \frac{Q(\rho)^{\omega}}{\varphi_\rho} - F_r
$$

The wholesale and retail demand are unchanged from previous sections.

The firm now has the following first-order conditions for its choice of retail and wholesale quantities

$$
\begin{aligned}
\frac{\partial \pi_{wr}}{\partial q_r} &= 0 = \frac{\sigma_r - 1}{\sigma_r}\left(\frac{q_r}{A_r}\right)^{-\frac{1}{\sigma_r}} - \omega c_r p_m \frac{Q(\rho)^{\omega-1}}{\varphi_\rho} \\
\frac{\partial \pi_{wr}}{\partial q_w} &= 0 = \frac{\sigma_w - 1}{\sigma_w}\left(\frac{q_w}{A_w}\right)^{-\frac{1}{\sigma_w}} - \omega c \Omega_w^{-\frac{1}{\alpha}}\left(\int_{\omega\in\Omega_w} p(\kappa)^{\alpha(1-\sigma_r)}\, d\omega\right)_w^{\frac{1}{\alpha}} p_m \frac{Q(\rho)^{\omega-1}}{\varphi_\rho}
\end{aligned}
$$

which imply

$$
\begin{aligned}
q_r &= A_r\left(\frac{\sigma_r - 1}{\sigma_r \omega}\frac{1}{c_r p_m}\frac{\varphi_\rho}{Q(\rho)^{\omega-1}}\right)^{\sigma_r} \\
q_w &= A_w\left(\frac{\sigma_w - 1}{\sigma_w \omega}\frac{1}{c_w p_m}\frac{\varphi_\rho}{Q(\rho)^{\omega-1}}\right)^{\sigma_w}
\end{aligned}
$$

which, combined with the definition of Q, provide an implicit solution for the firm's optimal decisions. Furthermore,

$$
\begin{aligned}
p_r(\rho) &= \frac{\sigma_r \omega}{\sigma_r - 1} c_r p_m \frac{Q(\rho)^{\omega-1}}{\varphi_\rho} \\
p_w(\rho) &= \frac{\sigma_w \omega}{\sigma_w - 1} c_w p_m \frac{Q(\rho)^{\omega-1}}{\varphi_\rho}
\end{aligned}
$$

for the Q implicitly defined by the first-order conditions.

Optimal selling rules. Firms enter retail, wholesale, or both when they (expect to) earn positive profits from doing so. We can describe these entry decisions in terms of cutoff rules for productivity. In particular, assuming $0 < F_r < F_w$,

$$0 < \tilde{\varphi}_r < \tilde{\varphi}_w < \tilde{\varphi}_{wr} \tag{3.3}$$

We note the following:

1. The first inequality ($0 < \tilde{\varphi}_r$) is straightforward: A retail firm with $\varphi = 0$ will earn zero variable profits and thus earn negative profits overall.
2. The second inequality is because switching from retail to wholesale implies increased fixed costs, and therefore in equilibrium must carry variable benefits (otherwise no firms would wholesale, which is impossible since then no firms could retail). These variable benefits are increasing in productivity and so we get the usual single crossing condition.
3. And the third inequality is because doing both implies a greater fixed cost than selling wholesale only. However, it offers variable benefits since the same quantity spread over wholesale and retail will earn greater profits than selling wholesale only. This variable benefit is strictly increasing in productivity and hence the higher cutoff.

Equilibrium price indices

First, the wholesale pricing index will be

$$\begin{aligned} P_w^{1-\sigma_w} &= \frac{1}{\Omega_w} \int\limits_{\varphi \in \Omega_w} p_w(\varphi)^{\frac{\sigma_w - 1}{\alpha(\sigma_r - 1)}} \, dG(\varphi) \\ &= \frac{1}{M\left(G(\tilde{\varphi}_{wr}) - G(\tilde{\varphi}_w)\right)} \int\limits_{\tilde{\varphi}_w}^{\tilde{\varphi}_{wr}} \hat{p}_w(\varphi)^{\frac{\sigma_w - 1}{\alpha(\sigma_r - 1)}} \, dG(\varphi) \\ &+ \frac{1}{M\left(1 - G(\tilde{\varphi}_{wr})\right)} \int\limits_{\tilde{\varphi}_{wr}}^{1} p_w(\varphi)^{\frac{\sigma_w - 1}{\alpha(\sigma_r - 1)}} \, dG(\varphi) \end{aligned}$$

and the wholesale profit index will be

$$\begin{aligned} \tilde{P}_w^{1-\sigma_w} &= \Omega_w^{\frac{1}{\alpha}} \left(\int\limits_{\varphi \in \Omega_w} p(\varphi)^{\alpha(1-\sigma_r)} \, dG(\varphi) \right)^{-\frac{1}{\alpha}} \\ &= M^{\frac{1}{\alpha}} \left(1 - G(\tilde{\varphi}_r)\right)^{\frac{1}{\alpha}} \left(\int\limits_{\tilde{\varphi}_r}^{\tilde{\varphi}_w} \hat{p}_r(\varphi)^{\alpha(1-\sigma_r)} \, dG(\varphi) + \int\limits_{\tilde{\varphi}_{wr}}^{1} p_r(\varphi)^{\alpha(1-\sigma_r)} \, dG(\varphi) \right)^{-\frac{1}{\alpha}} \end{aligned}$$

The retail price index will be

$$P_r^{1-\sigma_r} = \int_{\tilde{\varphi}_r}^{\tilde{\varphi}_w} \hat{p}_r(\varphi)^{1-\sigma_r}\, dG(\varphi) + \int_{\tilde{\varphi}_{wr}}^{1} p_r(\varphi)^{1-\sigma_r}\, dG(\varphi)$$

Note that all the cutoffs are defined by the usual zero-profit conditions. Expenditure on wholesale will be the expenditure on resale only, less resale markups.

Costs: Span of Control Problems

Setup

Turning our attention to the costs of scaling firms up, we present a simple span-of-control model micro-founded by principal-agent problems where the employer has to spend time to monitor employees' effort. There are two agents, the principal (owner) and the agent (worker). The principal does not observe the agent's effort, but can spend time to monitor the worker. The principal spends time on sourcing and monitoring. The principal's payoff, i.e. firm profit, comes from two sources: revenue from sourcing and worker's output, but the revenue function does not need to be additively separable.

The worker faces the following:

- The worker is offered a wage w.
- She chooses to exert effort or not. Assume there are only two levels of effort: 0 and 1.
  - Putting in positive effort incurs a cost e to the worker.
  - Putting in zero effort incurs no cost.
- The worker faces a probability $q(t_m)$ of being caught shirking, which is increasing in principal's time spent on monitoring t_m.
- There are N workers in total.

The principal (employer) faces the following:

- The principal spends finite time on two tasks :
  - **Monitoring (t_m)**: Increases $q(t_m)$, the probability of catching the worker shirking $e = 0$, where $q'(t_m) > 0$ and $q''(t_m) < 0$.
  - **Sourcing (t_s)**: Generates returns $R(t_s, \varphi)$, where φ is the productivity of the firm. We assume that

    $$\frac{\partial R(t_s,\varphi)}{\partial t_s} > 0, \frac{\partial^2 R(t_s,\varphi)}{\partial\varphi\partial t_s} > 0$$

    In other words, there is more to be gained from investing in sourcing at larger firms.

- Time constraint: $Nt_m + t_s = T$.
- Principal gains $\pi(e)$ from worker, where:

$$\pi(e) = \begin{cases} 0 & \text{if } e = 0, \\ g(\varphi) > 0 & \text{if } e > 0. \end{cases}$$

Here $g'(\varphi) > 0$, so more productive firms make workers more productive too.

Payoffs

Worker's Utility. The worker chooses to either exert effort or shirk.

1. Exerting effort:
$$U(1) = w - e$$
2. Shirking:
$$U(0) = (1 - q(t_m))w$$
since if the worker gets caught she gets paid zero.

Incentive Compatibility Constraint. The worker chooses effort if $U(1) \geq U(0)$:

$$w - e \geq (1 - q(t_m))w.$$

Rearranging:
$$e \leq q(t_m)w.$$

Thus, the wage must satisfy:
$$w \geq \frac{e}{q(t_m)}.$$

In equilibrium, $w = \frac{e}{q(t_m)}$.

Principal's Profit. The principal's total profit is:

$$\Pi = R(t_s, \varphi, N) - Nw \tag{3.4}$$

where $t_s = T - Nt_m$ and $R(\cdot)$ is a the revenue function that depends on the firm's productivity, owner's time spent on sourcing, and the number of workers. Conceptually, firm earns revenue from two sources: workers' effort (e.g., tending to customers, tidying up shops) and owner's time spent on sourcing (e.g., finding new customers and cheaper suppliers). These two parts may or may not enter the revenue function in an additively separable way.

Optimal time allocation. The principal chooses t_m (monitoring time) to maximize profit:
$$\max_{t_m} \ \Pi = R(t_s, \varphi, N) - Nw$$

Differentiate with respect to t_m yields the following first-order condition

$$\frac{\partial \Pi}{\partial t_m} = N\frac{eq'(t_m)}{q(t_m)^2} - \frac{\partial R(T - Nt_m, \varphi)}{\partial (T - Nt_m)} = 0 \tag{3.5}$$

The first part is the lower wages to be achieved by spending time on monitoring. The second part is the higher returns on sourcing that can be achieved by spending more time on monitoring.

What would an increase in the number of workers do? Suppose the firm with some fixed productivity level φ is already at optimum with choices (N^*, t_m^*). Then by the implicit function theorem,

$$\frac{\partial N}{\partial t_m} = -\frac{\frac{\partial F}{\partial N}(N, t_m(N))}{\frac{\partial F}{\partial t_m}(N, t_m(N))}$$

where $F(N, t_m) = N\frac{eq'(t_m)}{q(t_m)^2} - \frac{\partial R(T-Nt_m,\varphi)}{\partial (T-Nt_m)}$.

Leontief revenue function

We now specify the revenue function to be a Leontief where worker labor and owner's time spent on sourcing are perfect complements. Let

$$\Pi = g(\varphi)\min\{kN, T - Nt_m\} - N\frac{e}{q(t_m)}$$

where k is a scaling factor, $t_r = s = T - Nt_m$ is the owner's time spent on sourcing, and same as before $\frac{e}{q(t_m)}$ is the wage per worker.

Prediction 1: Higher productivity firms spend more time on sourcing.

If we only think about maximizing revenue, the owner will choose t_m^* such that $T - Nt_m^* = kN$, so

$$t_m^* = \frac{T}{N} - k$$

However, would the owner want to spend less time on sourcing and more time on monitoring? When $KN > T - Nt_m$,

$$\frac{\partial \Pi}{\partial t_m} = N\frac{eq'(t_m)}{q(t_m)^2} - g(\varphi)N = N\left(\frac{eq'(t_m)}{q(t_m)^2} - g(\varphi)\right)$$

Thus, for high φ firms, $\frac{\partial \Pi}{\partial t_m}$ always, so they will choose $t_m^* = \frac{T}{N} - k$ which maximizes revenue.

For low φ firms, $\frac{\partial \Pi}{\partial t_m} > 0$ for some values of t_m, so they will choose $t_m^* > \frac{T}{N} - k$. Intuitively, in the extreme case where $\varphi = 0$, the firm will spend all their time on monitoring to lower wages.

Prediction 2: Conditional on productivity, an increase in the number of workers *can* increase time on sourcing.

Now we fix the firm productivity φ. Increasing N creates two effects in the opposite direction:

1. It rewards more time spent on sourcing, as KN increases in the revenue function.
2. It rewards more time spent on monitoring, as there are now more workers to monitor and a higher amount of wages to pay.

First we consider the high-φ case where the firm is already choosing $t_m^* = \frac{T}{N} - k$. Suppose N increases by 1.

Case 1: Choose $t_s^* = k(N+1)$ to maximize revenue. Then $t_m^* = \frac{T}{N+1} - k$, and

$$\Pi_{N+1,1} = g(\varphi)k(N+1) - (N+1)\frac{e}{q(\frac{T}{N+1} - k)}$$

Case 2: Choose $t_s^* = kN$, so the time spent on sourcing remains unchanged. Then

$$t_m^* = \frac{T - kN}{N+1} = \frac{T}{N+1} - k\frac{N}{N+1}$$

and

$$\Pi_{N+1,2} = g(\varphi)kN - (N+1)\frac{e}{q(\frac{T}{N+1} - k\frac{N}{N+1})}$$

Thus

$$\Pi_{N+1,1} - \Pi_{N+1,2} = g(\varphi)k - (N+1)\left[\frac{e}{q(\frac{T}{N+1} - k)} - \frac{e}{q(\frac{T}{N+1} - k\frac{N}{N+1})}\right]$$

This is positive, i.e., keeping sourcing time fixed is not optimal when N increases by 1, if φ is high, N is low, or if $q'(t_m)$ is small. In other words, for lower number of initial employees and higher productivity firms, increasing the number of employees will induce the owner to spend more time on sourcing. Of course, this does not necessarily show that $t_s^* = k(N+1)$ is the optimal strategy. There could be some other value of t_s that achieves the maximum. But this shows that keeping t_s at the same value as before the firm size increases is not an optimal strategy. The derivation also shows that monitoring time per worker t_m^* decreases so the wage $w = e/q(t_m^*)$ increases, which is consistent with the empirical observation that larger firms pay a higher hourly wage.

Now consider a firm with low φ, i.e. those that already chose a $t_s^* < kN$ with N employees because for them the returns to sourcing is low. Then by the previous discussion

$$\frac{\partial \Pi}{\partial t_m} = N\left(\frac{eq'(t_m)}{q(t_m)^2} - g(\varphi)\right) > 0$$

for some $t_m > t_m^* = T/N - k$. If at N employees there is an interior solution, i.e. there is some t_m such that the above expression equals 0, then that value of t_m will remain unchanged when we increase N. In this case, sourcing time t_s decreases as there are more employees, but the owner still chooses to spend the same amount of time monitoring each employee.

Calibration

Once we complete the model, we aim to calibrate the following moments:

Parameter	Interpretation
E_r	Total expenditure by consumers
σ_r	Elasticity of substitution across retailers
α	Match dispersion parameter across wholesalers (part of σ_w)
ω	Cost elasticity (with regards to quantity aggregate Q)
c_r	Cost of a retail sale
c_w	Cost of a wholesale sale
a, b	Shapes of the productivity distribution
p_m	Price charged by manufacturer
F_r	Fixed cost entering retail
F_w	Fixed cost of entering wholesale
M	Measure of firms

Table 3.1: Parameters to calibrate

3.5 Conclusion

This paper presents an early version of both the empirical findings and the model that study the costs and benefits of the co-location of a large number of small retail firms in Lagos, Nigeria. We show that retail firms often source from very nearby wholesalers within their own market to avoid paying large fixed costs and take advantage of the scale economy offered by the wholesalers. We also document that labor and supervision costs are increasing in the number of employees, consistent with the presence of span of control problems. Conceptually, we expect that these two forces will lead to modest gains from turning small traders into employees at large retail companies. Future work will complete the full version of the model, calibrate model parameters, simulate the model structurally, and derive counterfactual predictions based on various policy changes, such as relaxing span of control problems or reducing the fixed cost of sourcing from the manufacturer.

3.6 Figures and Tables

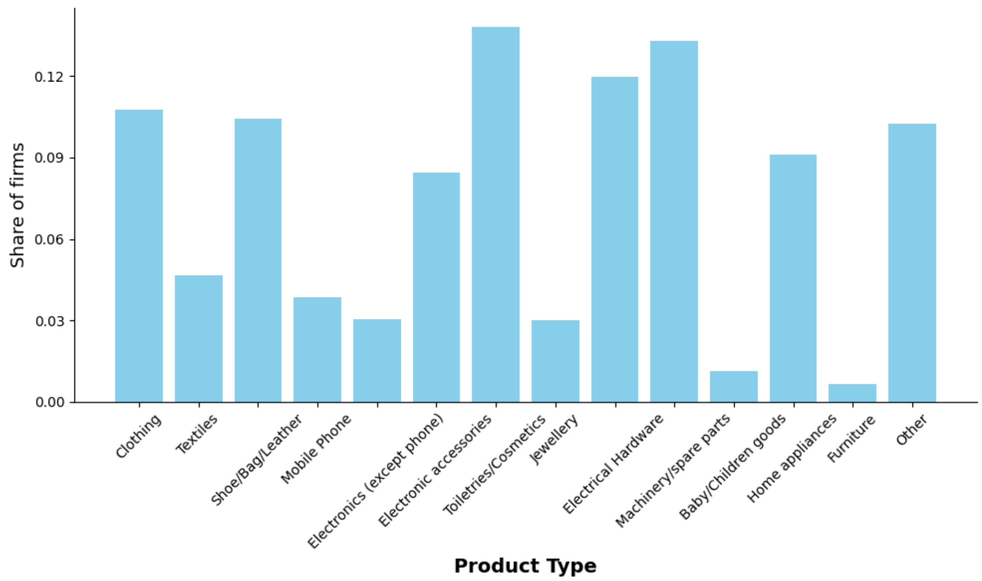


Figure 3.1: Distribution of firm product types

Note: Figure shows the share of firms in the LTS5 sample that sell goods in each product category. Shares do not necessarily add up to 1 because firms may sell products of multiple types.

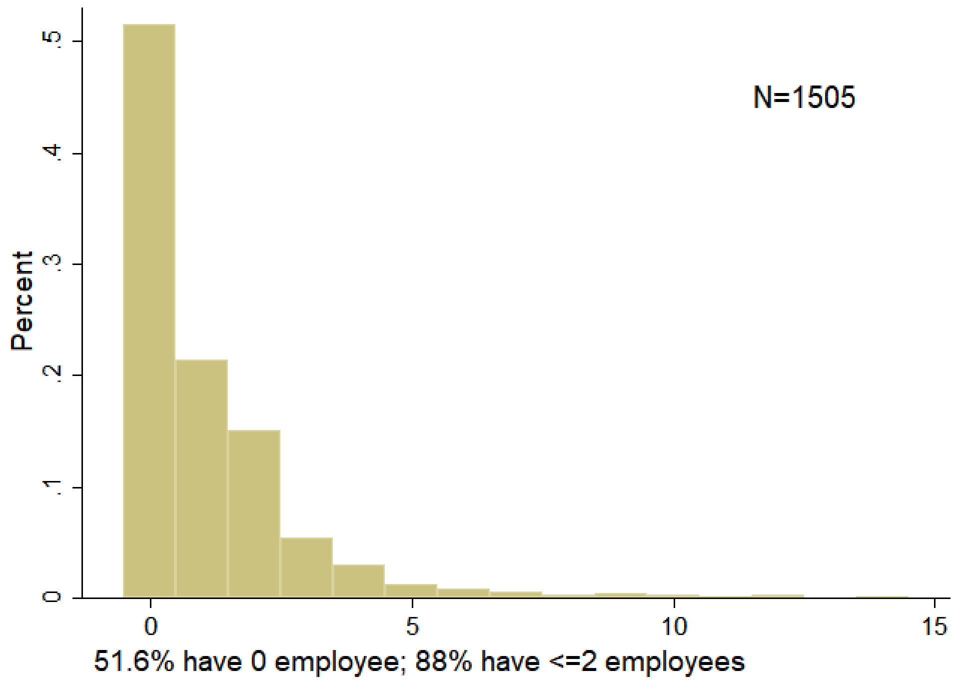


Figure 3.2: Distribution of firm size

Note: Figure shows the distribution of the number of employees per firm in the Lagos Trader SUrvey Round 5 sample.

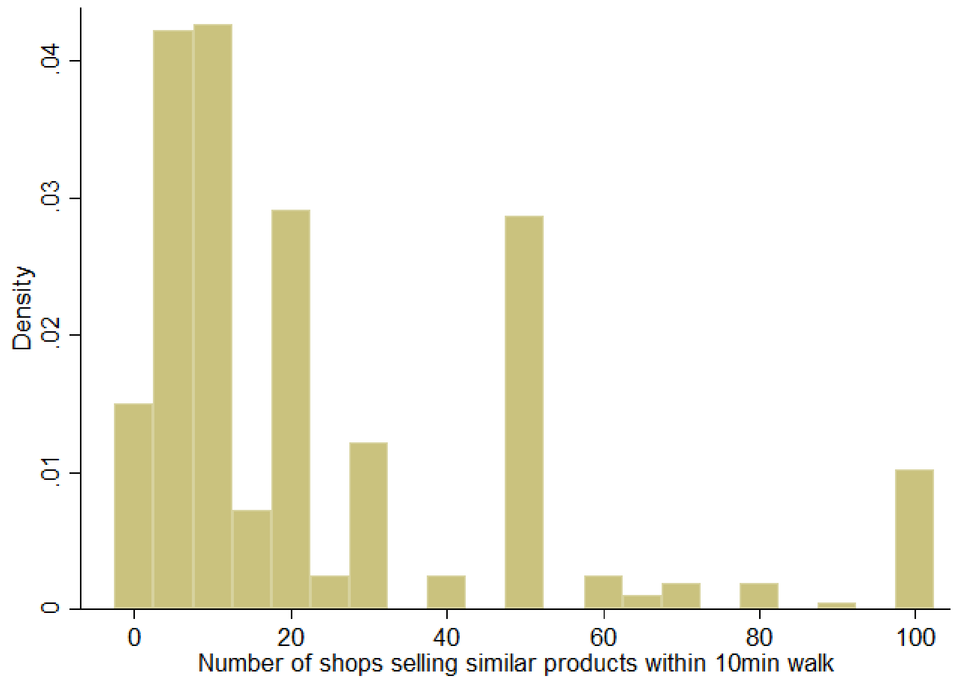


Figure 3.3: Distribution of number of shops in own market

Note: Figure shows the distribution of respondents' estimates of the number of shops in their own market selling the same or similar products.

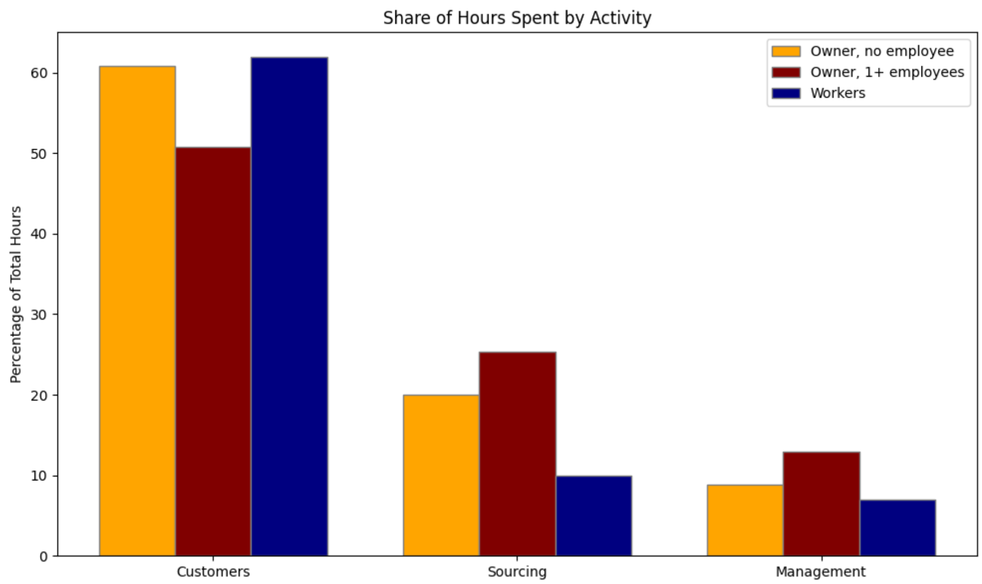
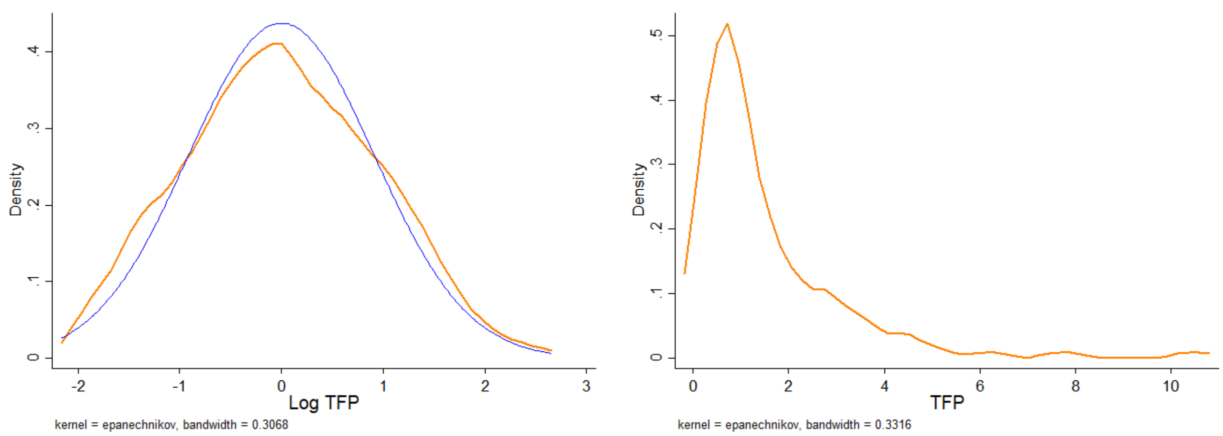


Figure 3.4: Summary of Time Use Patterns

Note: Figure shows the mean of the percentages of time spent on attending to customers, sourcing goods, and management for firm owners with zero employees, firm owners with one or more employees, and workers.



(a) Log TFP

(b) TFP

Figure 3.5: TFP Distributions

Note: Figure shows the kernel density estimates of the distribution of estimated log TFP and TFP. The blue line is the normal distribution.

Table 3.2: Summary statistics

	Wholesale	Retail	Both	Total
Number of employees	1.180	0.630	1.098	1.052
Total value of property (USD)	84845.9	12184.5	39214.2	42580.9
Annual revenue (USD)	56614.3	4769.5	26687.1	27478.9
Annual take-home profit (USD)	4735.3	1354.2	2726.3	2796.6
% of sales from retail	.	.	44.97	
Importer	0.695	0.160	0.537	0.513
Any sourcing within market	0.386	0.488	0.481	0.468
Any sale within market	0.830	0.476	0.787	0.751
Observations	200	181	1124	1505

Note: Table shows the mean of variables on the left-most column among firms in the sample that only do wholesale, only do retail, and do both wholesale and retail. The right-most column shows the mean for the entire sample. We use 1 USD = 1,000 NGN as the exchange rate.

Table 3.3: Collaboration among Traders

Activity	**Percentage**
Give or receive customer referrals	58%
Give or receive introduction to a supplier or agent	20%
Borrow or lend stocks	18%
Source products together on the same order	9%
Share shop space or stores	7%
Give or receive a gift or loan	21%
Share equipment	30%
Equipment shared among those that do:	
Generators	89%
Fans	13%
Calculators	5%
Observations	598

Note: Table shows the percentage of shop owners who engage in collaborative activities and share equipments with other businesses in the same market. The sample size is smaller than the full sample because only a randomly selected subset of respondents were asked the module containing those questions in order to reduce the length of the survey.

Table 3.4: Summary of Time Use Patterns

Owner, no employee:					
	mean	sd	p25	p50	p75
Attending to customers	60.83	16.61	50.00	60.00	70.00
Sourcing goods	20.07	12.28	10.00	20.00	30.00
Management	8.84	6.27	5.00	10.00	10.00
Owner, 1+ employees:					
Attending to customers	50.76	19.59	40.00	50.00	63.64
Sourcing goods	25.34	15.51	10.00	20.00	30.00
Management	12.92	9.99	9.09	10.00	20.00
Workers:					
Attending to customers	61.91	18.73	50.00	60.00	70.00
Sourcing goods	9.97	10.76	0.00	10.00	15.00
Management	6.93	7.99	0.00	5.00	10.00
Observations	606				

Note: All numbers are in percentages. Table shows the distribution of time spent on attending to customers, sourcing goods (including looking for new suppliers or products, making purchases, taking deliveries, etc.), and management (including record keeping, loans, hiring or supervising workers, etc.) The omitted category is "other tasks". The sample size is smaller than the full sample because only a randomly selected subset of respondents were asked the module containing those questions in order to reduce the length of the survey.

Table 3.5: Costs of Scaling Up

	Workers pay per hour (USD)	% hours on management	Sophisticated supervision	Any Training	Log Revenue per worker	Log Profit per worker
Number of employees	1.40**	0.97*	0.02**	0.96	-0.13	-0.15***
	(0.63)	(0.52)	(0.01)	(1.07)	(0.09)	(0.05)
N	246	293	729	295	115	287

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Table shows the coefficients of the variable on the top row on the number of employees. The dependent variable in Column 3 is an indicator for whether the owner engages in any supervision methods beyond watching the worker. The dependent variable in Column 3 is an indicator for whether the owner provides any training to their workers. All regressions control for whether the firm imports, an indicator for the survey round (original or new respondents), indicators of wholesale only/retail only/both, enumerator FEs and product category FEs. Sample only includes firms with one or more employees.

Table 3.6: Productivity Estimation

	(1) Log total revenue	(2) Log total revenue
Log capital	0.45***	0.34***
	(0.08)	(0.10)
Log labor (including the owner)	0.40*	0.14
	(0.20)	(0.24)
Log input	0.19**	0.28**
	(0.09)	(0.13)
N	114	114
Controls	No	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: Table shows the Total Factor Productivity (TFP) regression results. Column 2 controls for an indicator of whether the firm imports, an indicator for the survey round (original or new respondents), indicators of wholesale only/retail only/both, enumerator FEs and product category FEs.

Bibliography

Abebe, Girum, Margaret McMillan, and Michel Serafinelli (2022). "Foreign direct investment and knowledge diffusion in poor locations". en. In: *Journal of Development Economics* 158, p. 102926. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2022.102926. URL: https://www.sciencedirect.com/science/article/pii/S0304387822000827.

Acemoglu, Daron and David Autor (2011). "Skills, Tasks and Technologies: Implications for Employment and Earnings*". In: *Handbook of Labor Economics*. Ed. by David Card and Orley Ashenfelter. Vol. 4. Elsevier, pp. 1043–1171. DOI: 10.1016/S0169-7218(11)02410-5. URL: https://www.sciencedirect.com/science/article/pii/S0169721811024105.

Adão, Rodrigo, Michal Kolesár, and Eduardo Morales (2019). "Shift-Share Designs: Theory and Inference*". In: *The Quarterly Journal of Economics* 134.4, pp. 1949–2010. ISSN: 0033-5533. DOI: 10.1093/qje/qjz025. URL: https://doi.org/10.1093/qje/qjz025.

Agesa, Richard U., Jacqueline Agesa, and Andrew Dabalen (2013). "Sources of the Persistent Gender Wage Gap along the Unconditional Earnings Distribution: Findings from Kenya". In: *Oxford Development Studies* 41.1, pp. 76–103. ISSN: 1360-0818. DOI: 10.1080/13600818.2013.770304. URL: https://doi.org/10.1080/13600818.2013.770304.

Aker, Jenny C., Joshua Evan Blumenstock, and Brian Dillon (2020). *How Important is the Yellow Pages? Experimental Evidence from Tanzania*. en. SSRN Scholarly Paper. Rochester, NY. URL: https://papers.ssrn.com/abstract=3560296.

Alarakhia, Mehjabeen et al. (2023). *Why Women Earn Less: Gender Pay Gap and Labour-Market Inequalities in East and Southern Africa*. Tech. rep. UN Women. URL: https://africa.unwomen.org/sites/default/files/2023-10/gpg_regional_un_women_final.pdf.

Alderotti, Giammarco, Chiara Rapallini, and Silvio Traverso (2023). "The Big Five personality traits and earnings: A meta-analysis". In: *Journal of Economic Psychology* 94, p. 102570. ISSN: 0167-4870. DOI: 10.1016/j.joep.2022.102570. URL: https://www.sciencedirect.com/science/article/pii/S0167487022000812.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P Vasquez (2022). "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages*". In: *The Quarterly Journal of Economics* 137.3, pp. 1495–1552. ISSN: 0033-5533. DOI: 10.1093/qje/qjac006. URL: https://doi.org/10.1093/qje/qjac006.

Alwago, Wycliffe Obwori (2024). “Decomposing the Gender Wage Gap in the Urban Labor Market in Kenya”. en. In: *Studia Universitatis Vasile Goldis” Arad – Economics Series* 34.4, pp. 1–33. ISSN: 2285-3065. DOI: 10.2478/sues-2024-0016. URL: https://www.sciendo.com/article/10.2478/sues-2024-0016.

Anderson, Simon P., André De Palma, and Jacques-François Thisse (1987). “The CES is a discrete choice model?” In: *Economics Letters* 24.2, pp. 139–140. ISSN: 0165-1765. DOI: 10.1016/0165-1765(87)90239-4. URL: https://www.sciencedirect.com/science/article/pii/0165176587902394.

Appleton, Simon, John Hoddinott, and Pramila Krishnan (1999). “The Gender Wage Gap in Three African Countries”. In: *Economic Development and Cultural Change* 47.2. Publisher: The University of Chicago Press, pp. 289–312. ISSN: 0013-0079. DOI: 10.1086/452402. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/452402.

Aroca, Patricio, Mariano Bosch, and William F. Maloney (2005). “Spatial Dimensions of Trade Liberalization and Economic Convergence: Mexico 1985–2002”. In: *The World Bank Economic Review* 19.3, pp. 345–378. ISSN: 0258-6770. DOI: 10.1093/wber/lhi018. URL: https://doi.org/10.1093/wber/lhi018.

Atkin, David and Dave Donaldson (2015). *Who’s Getting Globalized? The Size and Implications of Intra-national Trade Costs*. Working Paper. DOI: 10.3386/w21439. URL: https://www.nber.org/papers/w21439.

Atkin, David, Benjamin Faber, and Marco Gonzalez-Navarro (2018). “Retail Globalization and Household Welfare: Evidence from Mexico”. In: *Journal of Political Economy* 126.1. Publisher: The University of Chicago Press, pp. 1–73. ISSN: 0022-3808. DOI: 10.1086/695476. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/695476.

Atkin, David, Amit K. Khandelwal, and Adam Osman (2017). “Exporting and Firm Performance: Evidence from a Randomized Experiment*”. In: *The Quarterly Journal of Economics* 132.2, pp. 551–615. ISSN: 0033-5533. DOI: 10.1093/qje/qjx002. URL: https://doi.org/10.1093/qje/qjx002.

Autor, David, David Dorn, and Gordon H. Hanson (2013). “The China Syndrome: Local Labor Market Effects of Import Competition in the United States”. en. In: *American Economic Review* 103.6, pp. 2121–2168. ISSN: 0002-8282. DOI: 10.1257/aer.103.6.2121. URL: https://www.aeaweb.org/articles?id=10.1257/aer.103.6.2121.

Autor, David H. and Michael J. Handel (2013). “Putting Tasks to the Test: Human Capital, Job Tasks, and Wages”. In: *Journal of Labor Economics* 31.2. Publisher: [The University of Chicago Press, Society of Labor Economists, NORC at the University of Chicago], S59–S96. ISSN: 0734-306X. DOI: 10.1086/669332. URL: https://www.jstor.org/stable/10.1086/669332.

Bach, Philipp, Victor Chernozhukov, and Martin Spindler (2024). “Heterogeneity in the US gender wage gap”. In: *Journal of the Royal Statistical Society Series A: Statistics in Society* 187.1, pp. 209–230. ISSN: 0964-1998. DOI: 10.1093/jrsssa/qnad091. URL: https://doi.org/10.1093/jrsssa/qnad091.

Banerjee, Abhijit et al. (2022). *Do Collusive Norms Maximize Profits? Evidence from a Vegetable Market Experiment in India*. en. Tech. rep. w30360. Cambridge, MA: National

Bureau of Economic Research, w30360. DOI: 10.3386/w30360. URL: http://www.nber.org/papers/w30360.pdf.

Bartelme, Dominick et al. (2019). *The Textbook Case for Industrial Policy: Theory Meets Data*. en. Tech. rep. w26193. Cambridge, MA: National Bureau of Economic Research, w26193. DOI: 10.3386/w26193. URL: http://www.nber.org/papers/w26193.pdf.

Bartik, Timothy J. (1991). *Who Benefits from State and Local Economic Development Policies?* W.E. Upjohn Institute. ISBN: 978-0-88099-113-1. URL: https://www.jstor.org/stable/j.ctvh4zh1q.

Bassi, Vittorio, Jung Hyuk Lee, et al. (2023). *Self-Employment Within the Firm*. Working Paper. DOI: 10.3386/w31740. URL: https://www.nber.org/papers/w31740.

Bassi, Vittorio, Raffaela Muoio, et al. (2022). "Achieving Scale Collectively". en. In: *Econometrica* 90.6, pp. 2937–2978. ISSN: 0012-9682. DOI: 10.3982/ECTA18773. URL: https://www.econometricsociety.org/doi/10.3982/ECTA18773.

Bastagli, Francesca et al. (2016). *Cash transfers: what does the evidence say?* en. Tech. rep. Overseas Development Institute.

Becerra, Oscar and José-Alberto Guerra (2023). "Personal safety first: Do workers value safer jobs?" In: *Journal of Economic Behavior & Organization* 212, pp. 996–1016. ISSN: 0167-2681. DOI: 10.1016/j.jebo.2023.06.017. URL: https://www.sciencedirect.com/science/article/pii/S0167268123002238.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on Treatment Effects after Selection among High-Dimensional Controls†". In: *The Review of Economic Studies* 81.2, pp. 608–650. ISSN: 0034-6527. DOI: 10.1093/restud/rdt044. URL: https://doi.org/10.1093/restud/rdt044.

Blanchard, Olivier and Lawrence Katz (1992). "Regional Evolutions". In: *Brookings Papers on Economic Activity* 23.1. Publisher: Economic Studies Program, The Brookings Institution, pp. 1–76. URL: https://econpapers.repec.org/article/binbpeajo/v_3a23_3ay_3a1992_3ai_3a1992-1_3ap_3a1-76.htm.

Blau, Francine D. and Lawrence M. Kahn (2017). "The Gender Wage Gap: Extent, Trends, and Explanations". en. In: *Journal of Economic Literature* 55.3, pp. 789–865. ISSN: 0022-0515. DOI: 10.1257/jel.20160995. URL: https://www.aeaweb.org/articles?id=10.1257/jel.20160995.

Blinder, Alan S. (1973). "Wage Discrimination: Reduced Form and Structural Estimates". In: *The Journal of Human Resources* 8.4. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], pp. 436–455. ISSN: 0022-166X. DOI: 10.2307/144855. URL: https://www.jstor.org/stable/144855.

Bloom, Nicholas et al. (2013). "Does Management Matter? Evidence from India *". In: *The Quarterly Journal of Economics* 128.1, pp. 1–51. ISSN: 0033-5533. DOI: 10.1093/qje/qjs044. URL: https://doi.org/10.1093/qje/qjs044.

Böheim, René and Philipp Stöllinger (2021). "Decomposition of the gender wage gap using the LASSO estimator". In: *Applied Economics Letters* 28.10. Publisher: Routledge _eprint: https://doi.org/10.1080/13504851.2020.1782332, pp. 817–828. ISSN: 1350-4851.

DOI: 10.1080/13504851.2020.1782332. URL: https://doi.org/10.1080/13504851.2020.1782332.

Boissiere, M., J. B. Knight, and R. H. Sabot (1985). "Earnings, Schooling, Ability, and Cognitive Skills". In: *The American Economic Review* 75.5. Publisher: American Economic Association, pp. 1016–1030. ISSN: 0002-8282. URL: https://www.jstor.org/stable/1818642.

Bonaccolto-Töpfer, Marina and Stephanie Briel (2022). "The gender pay gap revisited: Does machine learning offer new insights?" In: *Labour Economics* 78, p. 102223. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2022.102223. URL: https://www.sciencedirect.com/science/article/pii/S0927537122001130.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel (2022). "Quasi-Experimental Shift-Share Research Designs". In: *The Review of Economic Studies* 89.1, pp. 181–213. ISSN: 0034-6527. DOI: 10.1093/restud/rdab030. URL: https://doi.org/10.1093/restud/rdab030.

Buchmann, Nina, Carl Meyer, and Colin D. Sullivan (2024). "Paternalistic Discrimination". In: *mimeo*.

Bursztyn, Leonardo, Alessandra L. González, and David Yanagizawa-Drott (2020). "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia". en. In: *American Economic Review* 110.10, pp. 2997–3029. ISSN: 0002-8282. DOI: 10.1257/aer.20180975. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20180975.

Card, David (2001). "Immigrant Inflows, Native Outflows, and the Local Labor Market Impacts of Higher Immigration". In: *Journal of Labor Economics* 19.1. Publisher: The University of Chicago Press, pp. 22–64. ISSN: 0734-306X. DOI: 10.1086/209979. URL: https://www.journals.uchicago.edu/doi/10.1086/209979.

Card, David et al. (2018). "Firms and Labor Market Inequality: Evidence and Some Theory". In: *Journal of Labor Economics* 36.S1. Publisher: The University of Chicago Press, S13–S70. ISSN: 0734-306X. DOI: 10.1086/694153. URL: https://www.journals.uchicago.edu/doi/10.1086/694153.

Cawley, John, James Heckman, and Edward Vytlacil (2001). "Three observations on wages and measured cognitive ability". In: *Labour Economics* 8.4, pp. 419–442. ISSN: 0927-5371. DOI: 10.1016/S0927-5371(01)00039-2. URL: https://www.sciencedirect.com/science/article/pii/S0927537101000392.

Cheema, Ali et al. (2023). *Glass walls: Experimental evidence on access constraints faced by women*. Publication Title: Technical Report. URL: https://mgshss.lums.edu.pk/sites/default/files/inline-files/glass_walls_draft_020222.pdf.

Chen, Yutong et al. (2024). *Gender-Specific Transportation Costs and Female Time Use: Evidence from India's Pink Slip Program*. Working Paper. DOI: 10.3386/w32508. URL: https://www.nber.org/papers/w32508.

Chernozhukov, Victor, Denis Chetverikov, et al. (2016). *Double/Debiased Machine Learning for Treatment and Causal Parameters*. Version Number: 6. DOI: 10.48550/ARXIV.1608.00060. URL: https://arxiv.org/abs/1608.00060.

Chernozhukov, Victor, Denis Nekipelov, et al. (2018). *Plug-in regularized estimation of high dimensional parameters in nonlinear semiparametric models*. en. Tech. rep. The IFS. DOI: 10.1920/wp.cem.2018.4118. URL: https://www.ifs.org.uk/uploads/cemmap/wps/CWP411818.pdf.

Clarke, Paul and Annalivia Polselli (2023). *Double Machine Learning for Static Panel Models with Fixed Effects*. Version Number: 3. DOI: 10.48550/ARXIV.2312.08174. URL: https://arxiv.org/abs/2312.08174.

Crust, Erin E., Mary C. Daly, and Bart Hobijn (2020). *The Illusion of Wage Growth*. Tech. rep. 2020-26. FRBSF Economic Letter.

Cunha, Jesse M, Giacomo De Giorgi, and Seema Jayachandran (2019). "The Price Effects of Cash Versus In-Kind Transfers". In: *The Review of Economic Studies* 86.1, pp. 240–281. ISSN: 0034-6527. DOI: 10.1093/restud/rdy018. URL: https://doi.org/10.1093/restud/rdy018.

De La Rica, Sara, Lucas Gortazar, and Piotr Lewandowski (2020). "Job Tasks and Wages in Developed Countries: Evidence from PIAAC". In: *Labour Economics* 65, p. 101845. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2020.101845. URL: https://www.sciencedirect.com/science/article/pii/S092753712030049X.

De Loecker, Jan et al. (2016). "Prices, Markups, and Trade Reform". en. In: *Econometrica* 84.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/ECTA11042, pp. 445–510. ISSN: 1468-0262. DOI: 10.3982/ECTA11042. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA11042.

DellaVigna, Stefano and Matthew Gentzkow (2019). "Uniform Pricing in U.S. Retail Chains". In: *The Quarterly Journal of Economics* 134.4, pp. 2011–2084. ISSN: 0033-5533. DOI: 10.1093/qje/qjz019. URL: https://doi.org/10.1093/qje/qjz019.

Deming, David J. (2017). "The Growing Importance of Social Skills in the Labor Market*". In: *The Quarterly Journal of Economics* 132.4, pp. 1593–1640. ISSN: 0033-5533. DOI: 10.1093/qje/qjx022. URL: https://doi.org/10.1093/qje/qjx022.

Diamond, Rebecca (2016). "The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000". en. In: *American Economic Review* 106.3, pp. 479–524. ISSN: 0002-8282. DOI: 10.1257/aer.20131706. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20131706.

Diamond, Rebecca and Enrico Moretti (2021). *Where is Standard of Living the Highest? Local Prices and the Geography of Consumption*. Working Paper. DOI: 10.3386/w29533. URL: https://www.nber.org/papers/w29533.

Dicarlo, Emanuele et al. (2016). "The Skill Content of Occupations Across Low and Middle Income Countries: Evidence from Harmonized Data". en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2846315. URL: https://www.ssrn.com/abstract=2846315.

Dixit, Avinash K. and Joseph E. Stiglitz (1979). "Monopolistic Competition and Optimum Product Diversity: Reply". In: *The American Economic Review* 69.5. Publisher: American Economic Association, pp. 961–963. ISSN: 0002-8282. URL: https://www.jstor.org/stable/1813666.

Edin, Per-Anders et al. (2022). “The Rising Return to Noncognitive Skill”. en. In: *American Economic Journal: Applied Economics* 14.2, pp. 78–100. ISSN: 1945-7782. DOI: 10.1257/app.20190199. URL: https://www.aeaweb.org/articles?id=10.1257/app.20190199.

Egger, Dennis et al. (2022). “General Equilibrium Effects of Cash Transfers: Experimental Evidence From Kenya”. en. In: *Econometrica* 90.6, pp. 2603–2643. ISSN: 1468-0262. DOI: 10.3982/ECTA17945. URL: https://onlinelibrary.wiley.com/doi/abs/10.3982/ECTA17945.

Eurostat (2022). *Gender pay gap statistics - Statistics Explained*. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Gender_pay_gap_statistics.

Faber, Benjamin (2014). “Trade Liberalization, the Price of Quality, and Inequality: Evidence from Mexican Store Prices”. en. In.

Faber, Benjamin and Thibault Fally (2022). “Firm Heterogeneity in Consumption Baskets: Evidence from Home and Store Scanner Data”. In: *The Review of Economic Studies* 89.3, pp. 1420–1459. ISSN: 0034-6527. DOI: 10.1093/restud/rdab061. URL: https://doi.org/10.1093/restud/rdab061.

Fafchamps, Marcel et al. (2014). “Microenterprise growth and the flypaper effect: Evidence from a randomized experiment in Ghana”. In: *Journal of Development Economics* 106, pp. 211–226. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2013.09.010. URL: https://www.sciencedirect.com/science/article/pii/S0304387813001375.

Fajgelbaum, Pablo D, Eduardo Morales, et al. (2019). “State Taxes and Spatial Misallocation”. In: *The Review of Economic Studies* 86.1, pp. 333–376. ISSN: 0034-6527. DOI: 10.1093/restud/rdy050. URL: https://doi.org/10.1093/restud/rdy050.

Fajgelbaum, Pablo D. and Amit K. Khandelwal (2016). “Measuring the Unequal Gains from Trade *”. In: *The Quarterly Journal of Economics* 131.3, pp. 1113–1180. ISSN: 0033-5533. DOI: 10.1093/qje/qjw013. URL: https://doi.org/10.1093/qje/qjw013.

Feenstra, Robert C. (2003). “A homothetic utility function for monopolistic competition models, without constant price elasticity”. In: *Economics Letters* 78.1, pp. 79–86. ISSN: 0165-1765. DOI: 10.1016/S0165-1765(02)00181-7. URL: https://www.sciencedirect.com/science/article/pii/S0165176502001817.

Gagnon, Etienne (2009). “Price Setting during Low and High Inflation: Evidence from Mexico*”. In: *The Quarterly Journal of Economics* 124.3, pp. 1221–1263. ISSN: 0033-5533. DOI: 10.1162/qjec.2009.124.3.1221. URL: https://doi.org/10.1162/qjec.2009.124.3.1221.

— (2018). *Mexican Micro CPI Dataset: A User Guide*.

Goldin, Claudia (2014). “A Grand Gender Convergence: Its Last Chapter”. en. In: *American Economic Review* 104.4, pp. 1091–1119. ISSN: 0002-8282. DOI: 10.1257/aer.104.4.1091. URL: https://www.aeaweb.org/articles?id=10.1257/aer.104.4.1091.

Gopinath, Gita and Oleg Itskhoki (2010). “Frequency of Price Adjustment and Pass-Through”. In: *The Quarterly Journal of Economics* 125.2, pp. 675–727. ISSN: 0033-5533. DOI: 10.1162/qjec.2010.125.2.675. URL: https://doi.org/10.1162/qjec.2010.125.2.675.

Grant, Matthew and Meredith Startz (2024). “Cutting Out the Middleman: The Structure of Chains of Intermediation”. In.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti (2010). “Identifying Agglomeration Spillovers: Evidence from Winners and Losers of Large Plant Openings”. In: *Journal of Political Economy* 118.3. Publisher: The University of Chicago Press, pp. 536–598. ISSN: 0022-3808. DOI: 10.1086/653714. URL: https://www.journals.uchicago.edu/doi/full/10.1086/653714.

Guide, EconML User (2023). *Orthogonal/Double Machine Learning,* URL: https://econml.azurewebsites.net/spec/estimation/dml.html.

Hamory, Joan et al. (2021). “Twenty-year economic impacts of deworming”. In: *Proceedings of the National Academy of Sciences* 118.14. Publisher: Proceedings of the National Academy of Sciences, e2023185118. DOI: 10.1073/pnas.2023185118. URL: https://www.pnas.org/doi/10.1073/pnas.2023185118.

Hanushek, Eric A. and Ludger Woessmann (2008). “The Role of Cognitive Skills in Economic Development”. en. In: *Journal of Economic Literature* 46.3, pp. 607–668. ISSN: 0022-0515. DOI: 10.1257/jel.46.3.607. URL: https://www.aeaweb.org/articles?id=10.1257/jel.46.3.607.

Hardy, Morgan et al. (2024). “Allocating labor across small firms: Experimental evidence on information constraints”. In: *Journal of Development Economics* 171, p. 103345. ISSN: 0304-3878. DOI: 10.1016/j.jdeveco.2024.103345. URL: https://www.sciencedirect.com/science/article/pii/S0304387824000944.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006). “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior”. In: *Journal of Labor Economics* 24.3. Publisher: The University of Chicago Press, pp. 411–482. ISSN: 0734-306X. DOI: 10.1086/504455. URL: https://www.journals.uchicago.edu/doi/full/10.1086/504455.

Hermo, Santiago et al. (2022). “Labor Market Returns and the Evolution of Cognitive Skills: Theory and Evidence*”. In: *The Quarterly Journal of Economics* 137.4, pp. 2309–2361. ISSN: 0033-5533. DOI: 10.1093/qje/qjac022. URL: https://doi.org/10.1093/qje/qjac022.

Hornbeck, Richard and Enrico Moretti (2022). “Estimating Who Benefits from Productivity Growth: Local and Distant Effects of City Productivity Growth on Wages, Rents, and Inequality”. en. In: *The Review of Economics and Statistics*, pp. 1–49. ISSN: 0034-6535, 1530-9142. DOI: 10.1162/rest_a_01208. URL: https://direct.mit.edu/rest/article/doi/10.1162/rest_a_01208/111517/Estimating-Who-Benefits-from-Productivity-Growth.

Hsieh, Chang-Tai and Peter J. Klenow (2014). “The Life Cycle of Plants in India and Mexico *”. In: *The Quarterly Journal of Economics* 129.3, pp. 1035–1084. ISSN: 0033-5533. DOI: 10.1093/qje/qju014. URL: https://doi.org/10.1093/qje/qju014.

Hsieh, Chang-Tai and Benjamin A. Olken (2014). “The Missing ”Missing Middle””. en. In: *Journal of Economic Perspectives* 28.3, pp. 89–108. ISSN: 0895-3309. DOI: 10.1257/jep.28.3.89. URL: https://www.aeaweb.org/articles?id=10.1257/jep.28.3.89.

INEGI (2013). *INDICE NACIONAL DE PRECIOS AL CONSUMIDOR.* Tech. rep.

Jalota, Suhani and Lisa Ho (2024). *What Works For Her? How Work-from-Home Jobs Affect Female Labor Force Participation in Urban India*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4739387. URL: https://papers.ssrn.com/abstract=4739387.

Jaravel, Xavier (2019). "The Unequal Gains from Product Innovations: Evidence from the U.S. Retail Sector*". In: *The Quarterly Journal of Economics* 134.2, pp. 715–783. ISSN: 0033-5533. DOI: 10.1093/qje/qjy031. URL: https://doi.org/10.1093/qje/qjy031.

— (2021). "Inflation Inequality: Measurement, Causes, and Policy Implications". In: *Annual Review of Economics* 13.1. _eprint: https://doi.org/10.1146/annurev-economics-091520-082042, pp. 599–629. DOI: 10.1146/annurev-economics-091520-082042. URL: https://doi.org/10.1146/annurev-economics-091520-082042.

Jayachandran, Seema (2021). "Social Norms as a Barrier to Women's Employment in Developing Countries". en. In: *IMF Economic Review* 69.3, pp. 576–595. ISSN: 2041-417X. DOI: 10.1057/s41308-021-00140-w. URL: https://doi.org/10.1057/s41308-021-00140-w.

Jensen, Robert and Nolan H. Miller (2018). "Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India". en. In: *American Economic Review* 108.12, pp. 3583–3625. ISSN: 0002-8282. DOI: 10.1257/aer.20161965. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20161965.

Karasek, R. et al. (1998). "The Job Content Questionnaire (JCQ): an instrument for internationally comparative assessments of psychosocial job characteristics". eng. In: *Journal of Occupational Health Psychology* 3.4, pp. 322–355. ISSN: 1076-8998. DOI: 10.1037//1076-8998.3.4.322.

Kato, Atsuyuki (2012). "Productivity, returns to scale and product differentiation in the retail trade industry: an empirical analysis using Japanese firm-level data". In: *Journal of Productivity Analysis* 38.3. Publisher: Springer, pp. 345–353. ISSN: 0895-562X. URL: https://www.jstor.org/stable/43549957.

Klasen, Stephan (2019). "What Explains Uneven Female Labor Force Participation Levels and Trends in Developing Countries?" In: *The World Bank Research Observer* 34.2, pp. 161–197. ISSN: 0257-3032. DOI: 10.1093/wbro/lkz005. URL: https://doi.org/10.1093/wbro/lkz005.

Klier, Thomas and James Rubenstein (2017). "Mexico's Growing Role in the Auto Industry Under NAFTA: Who Makes What and What Goes Where - Federal Reserve Bank of Chicago". en. In: *Economic Perspectives* 41.6. URL: https://www.chicagofed.org/publications/economic-perspectives/2017/6.

Kline, Patrick and Enrico Moretti (2014). "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority *". In: *The Quarterly Journal of Economics* 129.1, pp. 275–331. ISSN: 0033-5533. DOI: 10.1093/qje/qjt034. URL: https://doi.org/10.1093/qje/qjt034.

Krugman, Paul R. (1979). "Increasing returns, monopolistic competition, and international trade". In: *Journal of International Economics* 9.4, pp. 469–479. ISSN: 0022-1996. DOI:

10.1016/0022-1996(79)90017-5. URL: https://www.sciencedirect.com/science/article/pii/0022199679900175.

Kunze, Astrid (2018). “The Gender Wage Gap in Developed Countries”. In: *The Oxford Handbook of Women and the Economy*. Ed. by Susan L. Averett, Laura M. Argys, and Saul D. Hoffman. Oxford University Press, p. 0. ISBN: 978-0-19-062896-3. DOI: 10.1093/oxfordhb/9780190628963.013.11. URL: https://doi.org/10.1093/oxfordhb/9780190628963.013.11.

Labor, U.S. Department of (2024). *What You Need to Know About the Gender Wage Gap — U.S. Department of Labor Blog*. URL: https://blog.dol.gov/2024/03/12/what-you-need-to-know-about-the-gender-wage-gap.

Lagakos, David (2016). “Explaining Cross-Country Productivity Differences in Retail Trade”. In: *Journal of Political Economy* 124.2. Publisher: The University of Chicago Press, pp. 579–620. ISSN: 0022-3808. DOI: 10.1086/684718. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/684718.

Lamadon, Thibaut et al. (2024). *Labor Market Matching, Wages, and Amenities*. en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.4895416. URL: https://papers.ssrn.com/abstract=4895416.

Lin, Dajun, Randall Lutter, and Christopher J. Ruhm (2018). “Cognitive performance and labour market outcomes”. In: *Labour Economics* 51, pp. 121–135. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2017.12.008. URL: https://www.sciencedirect.com/science/article/pii/S0927537117303329.

Lo Bello, Salvatore, Maria Laura Sanchez Puerta, and Hernan Winkler (2019). “From Ghana to America: The Skill Content of Jobs and Economic Development”. en. In: *IZA Discussion Paper Series*.

Marshall, Alfred (1920). *Principles of Economics: An Introductory Volume*. 8th. London: Macmillan.

McKelway, Madeline (2019). “Experimental evidence on the effects of women’s employment”. In: *Working Paper*. URL: https://scholar.google.com/scholar?cluster=15490416492627104447&hl=en&oi=scholarr.

Mel, Suresh de, David McKenzie, and Christopher Woodruff (2008). “Returns to Capital in Microenterprises: Evidence from a Field Experiment*”. In: *The Quarterly Journal of Economics* 123.4, pp. 1329–1372. ISSN: 0033-5533. DOI: 10.1162/qjec.2008.123.4.1329. URL: https://doi.org/10.1162/qjec.2008.123.4.1329.

— (2009). “Are Women More Credit Constrained? Experimental Evidence on Gender and Microenterprise Returns”. en. In: *American Economic Journal: Applied Economics* 1.3, pp. 1–32. ISSN: 1945-7782. DOI: 10.1257/app.1.3.1. URL: https://www.aeaweb.org/articles?id=10.1257/app.1.3.1.

— (2011). “Getting Credit to High Return Microentrepreneurs: The Results of an Information Intervention”. In: *The World Bank Economic Review* 25.3, pp. 456–485. ISSN: 0258-6770. DOI: 10.1093/wber/lhr023. URL: https://doi.org/10.1093/wber/lhr023.

Molnar, Christoph (2023). *Interpreting Machine Learning Models With SHAP: A Guide With Python Examples And Theory On Shapley Values*. URL: https://leanpub.com/shap.

Moretti, Enrico (2013). "Real Wage Inequality". en. In: *American Economic Journal: Applied Economics* 5.1, pp. 65–103. ISSN: 1945-7782. DOI: 10.1257/app.5.1.65. URL: https://www.aeaweb.org/articles?id=10.1257/app.5.1.65.

Moulton, Brent R. et al. (1997). "Addressing the Quality Change Issue in the Consumer Price Index". In: *Brookings Papers on Economic Activity* 1997.1. Publisher: Brookings Institution Press, pp. 305–366. ISSN: 0007-2303. DOI: 10.2307/2534705. URL: https://www.jstor.org/stable/2534705.

Mrázová, Monika and J. Peter Neary (2017). "Not So Demanding: Demand Structure and Firm Behavior". en. In: *American Economic Review* 107.12, pp. 3835–3874. ISSN: 0002-8282. DOI: 10.1257/aer.20160175. URL: https://www.aeaweb.org/articles?id=10.1257/aer.20160175.

Nordhaus, William D. (1998). "Quality Change in Price Indexes". en. In: *Journal of Economic Perspectives* 12.1, pp. 59–68. ISSN: 0895-3309. DOI: 10.1257/jep.12.1.59. URL: https://www.aeaweb.org/articles?id=10.1257/jep.12.1.59.

Nordman, Christophe J., Anne-Sophie Robilliard, and François Roubaud (2011). "Gender and ethnic earnings gaps in seven West African cities". In: *Labour Economics*. Labour markets in developing countries 18, S132–S145. ISSN: 0927-5371. DOI: 10.1016/j.labeco.2011.09.003. URL: https://www.sciencedirect.com/science/article/pii/S0927537111001035.

Nordman, Christophe Jalil, Leopold R. Sarr, and Smriti Sharma (2015). "Cognitive, Non-Cognitive Skills and Gender Wage Gaps: Evidence from Linked Employer-Employee Data in Bangladesh". en. In: *SSRN Electronic Journal*. ISSN: 1556-5068. DOI: 10.2139/ssrn.2655081. URL: https://www.ssrn.com/abstract=2655081.

Notowidigdo, Matthew J. (2020). "The Incidence of Local Labor Demand Shocks". In: *Journal of Labor Economics* 38.3. Publisher: The University of Chicago Press, pp. 687–725. ISSN: 0734-306X. DOI: 10.1086/706048. URL: https://www.journals.uchicago.edu/doi/full/10.1086/706048.

Oaxaca, Ronald (1973). "Male-Female Wage Differentials in Urban Labor Markets". In: *International Economic Review* 14.3. Publisher: [Economics Department of the University of Pennsylvania, Wiley, Institute of Social and Economic Research, Osaka University], pp. 693–709. ISSN: 0020-6598. DOI: 10.2307/2525981. URL: https://www.jstor.org/stable/2525981.

Oaxaca, Ronald L and Michael R Ransom (1994). "On discrimination and the decomposition of wage differentials". In: *Journal of Econometrics* 61.1, pp. 5–21. ISSN: 0304-4076. DOI: 10.1016/0304-4076(94)90074-4. URL: https://www.sciencedirect.com/science/article/pii/0304407694900744.

Ofer, Gur (1973). "Returns to Scale in Retail Trade*". en. In: *Review of Income and Wealth* 19.4, pp. 363–384. ISSN: 1475-4991. DOI: 10.1111/j.1475-4991.1973.tb00897.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-4991.1973.tb00897.x.

Randall, Laura, ed. (2015). *Changing structure of Mexico: political, social, and economic prospects*. 2nd ed. Columbia University seminar series. OCLC: ocm57208898. Armonk, N.Y: M.E. Sharpe. ISBN: 978-0-7656-1404-9 978-0-7656-1405-6.

Siddique, Zahra (2022). "Media-Reported Violence and Female Labor Supply". In: *Economic Development and Cultural Change* 70.4. Publisher: The University of Chicago Press, pp. 1337–1365. ISSN: 0013-0079. DOI: 10.1086/714009. URL: https://www.journals.uchicago.edu/doi/10.1086/714009.

Simonovska, Ina (2015). "Income Differences and Prices of Tradables: Insights from an Online Retailer". In: *The Review of Economic Studies* 82.4, pp. 1612–1656. ISSN: 0034-6527. DOI: 10.1093/restud/rdv015. URL: https://doi.org/10.1093/restud/rdv015.

Sockin, Jason (2021). *Show Me the Amenity: Are Higher-Paying Firms Better All Around?* en. SSRN Scholarly Paper. Rochester, NY. DOI: 10.2139/ssrn.3957002. URL: https://papers.ssrn.com/abstract=3957002.

Startz, Meredith (2024). "The value of face-to-face: Search and contracting problems in Nigerian trade". In.

Vafa, Keyon, Susan Athey, and David Blei (2023). *Decomposing Changes in the Gender Wage Gap over Worker Careers*.

Van Den Broeck, Goedele, Talip Kilic, and Janneke Pieters (2023). "Structural transformation and the gender pay gap in Sub-Saharan Africa". en. In: *PLOS ONE* 18.4. Ed. by José Alberto Molina, e0278188. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0278188. URL: https://dx.plos.org/10.1371/journal.pone.0278188.

Verick, Sher (2014). "Female labor force participation in developing countries". In: *IZA World of Labor*. ISSN: 20549571. DOI: 10.15185/izawol.87. URL: https://wol.iza.org/articles/female-labor-force-participation-in-developing-countries/v1.

Vitali, Anna (2024). "Consumer Search and Firm Location: Theory and Evidence from the Garment Sector in Uganda". en. In.

Walker, Michael W. et al. (2023). *Intergenerational Child Mortality Impacts of Deworming: Experimental Evidence from Two Decades of the Kenya Life Panel Survey*. Working Paper. DOI: 10.3386/w31162. URL: https://www.nber.org/papers/w31162.

Weidmann, Ben and David J. Deming (2021). "Team Players: How Social Skills Improve Team Performance". en. In: *Econometrica* 89.6, pp. 2637–2657. ISSN: 1468-0262. DOI: 10.3982/ECTA18461.

Yagan, Danny (2019). "Employment Hysteresis from the Great Recession". In: *Journal of Political Economy* 127.5. Publisher: The University of Chicago Press, pp. 2505–2558. ISSN: 0022-3808. DOI: 10.1086/701809. URL: https://www.journals.uchicago.edu/doi/abs/10.1086/701809.

Zhelobodko, Evgeny et al. (2012). "Monopolistic Competition: Beyond the Constant Elasticity of Substitution". en. In: *Econometrica* 80.6, pp. 2765–2784. ISSN: 1468-0262. DOI: 10.3982/ECTA9986.

Appendix A

Additional Materials for Chapter 1

A.1 Data appendix

Additional details on CPI microdata

This section documents additional details on the CPI microdata. It relies on prior documentation by Gagnon, 2009, Gagnon, 2018, Faber, 2014, and Atkin, Faber, and Gonzalez-Navarro, 2018. The sampling framework of the Mexican CPI microdata is described in detail in Banco de México (2002) and INEGI, 2013, and follows guidelines used by the United States that follow the IMF/ILO data collection standards for consumer price inflation measurement and reporting. The Mexican CPI is representative of cities in Mexico with over 20,000 inhabitants in the year 2000.

Each month, enumerators procure price quotes for approximately 83,500 specific goods and services obtained at approximately 21,000 establishments, across 46 metropolitan areas and 128 urban municipalities. In constructing of the Consumer Price Index, Banco de México/INEGI takes a simple average of prices across items in the lowest nest (within a product category) across varieties and locations. As a result, geographic coverage of the price data is designed to be representative of the urban Mexican economy. Price quotes are collected four times per month for food items and twice per month for other items. Then Banco de México/INEGI averages the price quotes within a month for each item and publishes the average price in the official government gazette, *Diario Oficial de la Federación*.

The CPI price series experienced two basket revisions: one in July 2002 and another in January 2011. Unfortunately, item keys before and after item updates differ, and a crosswalk for constructing price series across these periods is unavailable. Discussions with INEGI staff suggest that the crosswalk is likely permanently lost. The price data used in this study from 1993 to June 2002 use weights from the 1989 ENIGH survey, data from July 2002 to December 2010 use weights from the 2000 ENIGH survey, and data from January 2011 to July 2018 use weights from the 2008 ENIGH survey. ENIGH provides household expenditure shares across all 315 product categories, based on a nationally representative sample.

Price quotes include VAT if applicable (food and medicines are exempt from this tax) and are inclusive of any sales or promotions as long as they do not require the use of a coupon the consumer obtains elsewhere (e.g., in a newspaper cutout), a minimum purchase amount (e.g., buy three get one free promotions), or a specific type of payment (e.g., 10% off if the payment is with a Banamex credit card). Enumerators are trained not to simply transcribe posted prices. They instead inquire about the prices being actually charged, which is especially relevant in contexts in which negotiation is expected, such as at street stalls.

The CPI Catalogue of Generic Items gives every enumerator a complete description of the specific goods, points of purchase, and number of quotes for specific goods that are to be obtained. For example, in the CPI Catalogue for 2005, Tomato (a CPI product category) consists of Roma Tomato and Round Tomato varieties (specific goods). The enumerator obtains four quotes for the first and one for the latter. This is done so that the number of quotes reflects the relative importance of each specific good in the product category. Continuing with the example, in the city of Iguala, Tomato quotes are to be obtained as follows: one quote at a traditional market, one at a fruit and vegetable traditional retailer, two at Walmart, and one at a Tianguis (street market) (Atkin, Faber, and Gonzalez-Navarro, 2018). This is done so that price quotes reflect ENIGH consumption patterns across different kinds of retail outlets. As such, the price quotes collected are representative with respect to product varieties and retail channels, and capture what a representative Mexican consumer would face.

Additional details on ENIGH survey data

The National Household Income and Expenditure Survey (*Encuesta Nacional de Ingresos y Gastos de los Hogares*, ENIGH) is conducted every two years to offer a statistical overview of the behavior of household income and expenditure in terms of amount, source, and distribution. In the welfare quantification exercise, I use the 2000 ENIGH survey, which supplies the weights used in construction of the CPI from 2002 to 2010. The 2000 ENIGH survey sample contains 10,188 households in 43 out of the 46 metropolitan areas in which price quotes are collected.

Each household was administered a basic questionnaire and asked to fill out an expenditure notebook (*Cuadernillo de Gastos*), which asked the household details on the quantity, unit, place of purchase, and amount of specific product categories over a recall period (7 days for food and beverages at home and away from home, 1 month for cleaning items, personal care, education, culture, communication, and housing, 3 months for clothing and healthcare, and 6 months for household appliances, electronics, transportation, and other costs). Expenditure amounts are standardized into a three-month measure for each item in the publicly available data. Both ENIGH and CPI are based on the same product groups, which renders the matching between them in the welfare quantification exercise straightforward. ENIGH also asks households their monthly income, which I use to construct income quintiles.

A.2 Theory appendix

A micro-foundation for decreasing markups with increasing market size

In this section, I follow the general monopolistic competition model of Zhelobodko et al., 2012 and additional work by Jaravel, 2019 to illustrate a simple model in which an increase in product demand from increasing market size can lead to decreasing markups. The intuition is that an increase in market size leads to a higher rate of product entry, which leads to a higher number of substitutes for each existing variety. This exerts downward pressure on the prices of those existing varieties.

The model features L consumers with additively separable preferences over varieties:

$$\max_{x_i > 0} U = \int_0^N u(x_i)\ di, \text{ s.t. } \int_0^N p_i x_i\ di = E$$

Differentiating the Lagrangian with respect to x_i yields the inverse demand function

$$p_i = \frac{u'(x_i)}{\lambda}$$

where the Lagrange multiplier λ is

$$\lambda = \frac{\int_0^N x_i u'(x_i) di}{E}$$

Total quantity demanded is $q_i = Lx_i$. Each firm produces a single variety. Firms face the same cost function with fixed cost $F > 0$ and variable cost $V(q)$. The producer thus takes the residual demand curve as given and solves

$$\max \pi(q_i) = R(q_i) - C(q_i) = \frac{u'(x_i)}{\lambda} q_i - V(q_i) - F$$

The optimal markup of the producer is therefore

$$M_i^* = -\frac{x_i u''(x_i)}{u'(x_i)}$$

where the expression on the right-hand side is the elasticity of demand.

At the free-entry equilibrium, $\pi^*(q_i) = 0$ and labor market clearing implies

$$N^* C(q_i) = LE$$

Zhelobodko et al., 2012 derive comparative statics following an increase in market size, i.e., an increase in L:

$$\frac{dN^*}{dL} > 0,\ \frac{dx_i^*}{dL} < 0$$

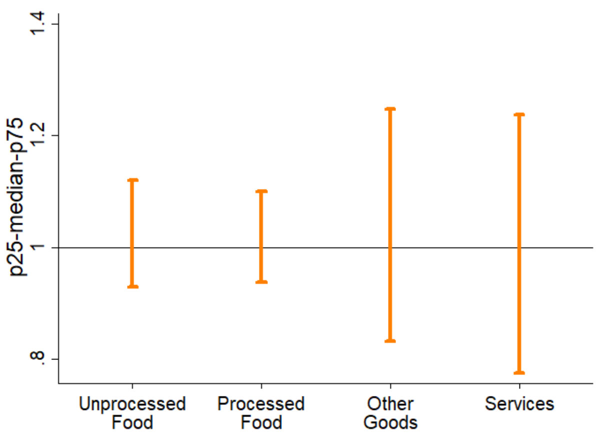
The equilibrium number of varieties N^* increases, because the higher demand encourages entry. Consumption on each variety per capita x_i^* decreases because in this case, in a market that provides more variety, consumers trade lower consumption of each variety for a more diversified basket of varieties. Importantly, these results do not depend on the shape of the cost function $V(\cdot)$.

Markup M_i^* may increase, decrease, or stay constant, depending on the curvature of the utility function. Under constant elasticity of substitution (CES) preferences, the price elasticity of demand for a given variety is constant, and thus the markup will not change following a fall in x_i^*. Conversely, if the price elasticity of demand increases when x_i^* decreases, the markup will fall, leading to pro-competitive effects of falling prices from an increase in market size. As a result, the markup and price increase when the elasticity of demand ϵ_i is decreasing with respect to quantity: $\frac{\partial \epsilon_i}{\partial x_i} < 0$, and decrease when $\frac{\partial \epsilon_i}{\partial x_i} > 0$.

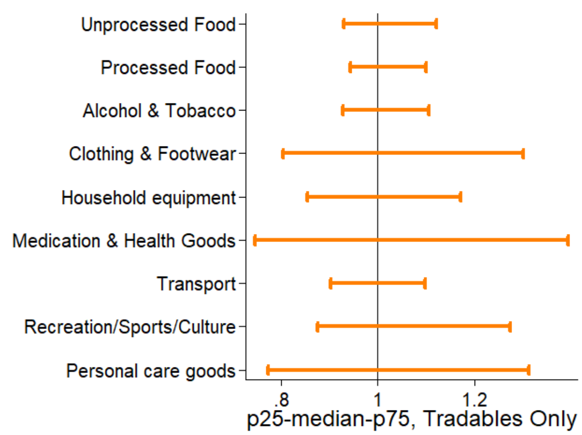
A.3 Additional figures and tables

Figure A.1: Spatial Dispersion of Prices, 1993 and 2018

October 1993:

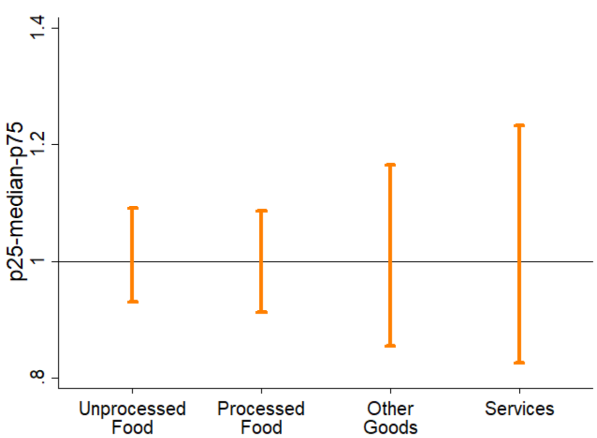


(a) Across four broad groups

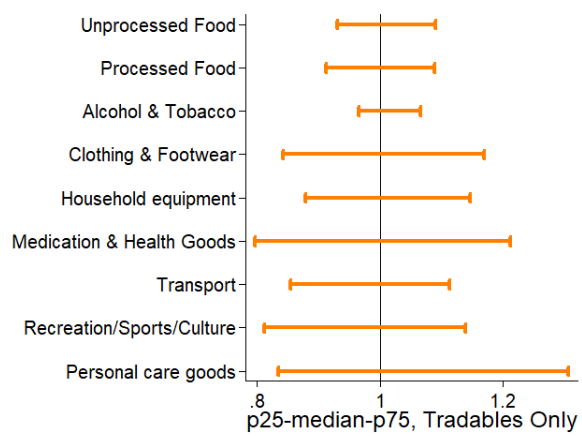


(b) Tradables only, across finer COICOP groups

March 2018:



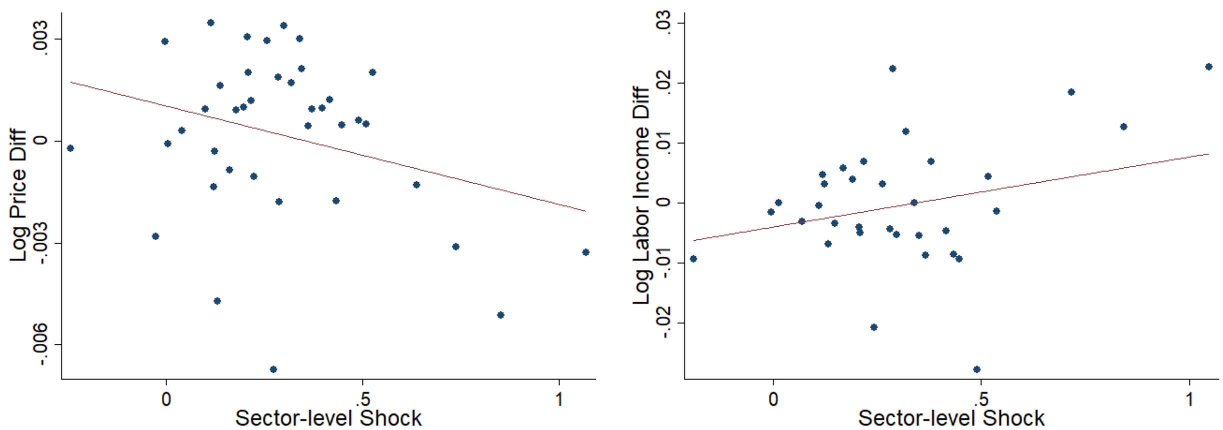
(c) Across four broad groups



(d) Tradables only, across finer COICOP groups

Note: Figures (a) and (c) show the interquartile range (p25-p75) of average prices across metropolitan areas for four broad categories of goods and services that partition the CPI data relative to the median, which is standardized to 1. Figures (b) and (d) show the interquartile range for 8 product groups within tradables that follow the Classification of Individual Consumption According to Purpose (COICOP).

Figure A.2: Reduced-form and first-stage relationships



(a) Reduced Form (b) First stage

Note: Figures show the binned scatter plots of reduced-form and first-stage relationships at shock level, where industry shocks are residualized on period indicators and weighted by their average exposure share following the procedure in Borusyak, Hull, and Jaravel, 2022.

Table A.1: Top 5 industries with the largest employment increases by period

1993-98		2003-08		2013-18	
Code	Sector	Code	Sector	Code	Sector
561	Administrative and Support Services	561	Administrative and Support Services	722	Food Services and Drinking Places
336	Transportation Equipment Manufacturing	722	Food Services and Drinking Places	336	Transportation Equipment Manufacturing
315	Apparel Manufacturing	461	Food retail	561	Administrative and Support Services
311	Food Processing	811	Repair and Maintenance	541	Professional, Scientific, and Technical Services
434	Wholesale trade of agricultural raw materials	462	Retail trade in self-service and department stores	311	Food Processing

Note: Tbale shows, in descending order of employment increases, the top five sector for each period of the analysis data along with their 3-digit SCIAN (*Sistema de Clasificación Industrial de América del Norte*, the Mexican equivalent of NAICS) code.

Table A.2: Main regressions, including region-by-period FEs

Reduced-form regression:								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Predicted local labor demand)	-0.0437*	0.133***	-0.0874***	0.00602	-0.122***	0.0102	-0.0955***	0.0703
	(0.0238)	(0.0391)	(0.0257)	(0.0362)	(0.0275)	(0.0416)	(0.0282)	(0.0519)
IV regression:								
Δ (Log wage bill per worker)	-0.122**	0.404**	-0.235***	0.0181	-0.230***	0.0208	-0.204***	0.148
	(0.0576)	(0.185)	(0.0657)	(0.111)	(0.0604)	(0.0859)	(0.0718)	(0.126)
Observations	85,867	3,445	64,373	18,049	64,373	18,049	64,373	18,049
IV F-stat	12.22	6.925	13.87	8.802	21.34	16.63	17.57	14.95
Controls					X	X	X	X
Leave-One-Out							X	X

Note: See Table 1.2 for detailed notes, all of which apply here as well. All regressions control for region-by-time-period fixed effects, in which the seven regions are defined according to the common definition employed in INEGI surveys: Northeast, Northwest, Southwest, Southeast, West, East, and Central South.

Table A.3: Robustness: Using log wage per adult, IV regression

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Log wage bill per worker)	-0.0192	1.072	-0.193***	0.261	-0.169***	0.105	-0.151***	0.171
	(0.106)	(0.757)	(0.0660)	(0.162)	(0.0452)	(0.0820)	(0.0517)	(0.118)
Observations	85,867	3,445	64,373	18,049	64,373	18,049	64,373	18,049
IV F-stat	10.50	2.435	9.829	11.37	11.48	10.83	11.61	13.13
Controls					X	X	X	X
Leave-One-Out							X	X

Note: See Table 1.2 for detailed notes, all of which apply here as well. The IV endogeneous variable is the difference in log wage per adult, interpolated from the decennial Census.

Table A.4: Excluding municipalities in northern border states

Reduced-form regression:								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Predicted local labor demand)	-0.0278	0.0676	-0.0730**	0.0621**	-0.118***	0.0650	-0.116***	0.0757
	(0.0297)	(0.0416)	(0.0318)	(0.0289)	(0.0245)	(0.0398)	(0.0218)	(0.0532)
IV regression:								
Δ (Log wage bill per worker)	-0.105	6.910	-0.237***	0.311	-0.208***	0.125	-0.221***	0.142
	(0.0741)	(113.4)	(0.0549)	(0.283)	(0.0401)	(0.111)	(0.0479)	(0.136)
Observations	66,377	2,598	49,920	13,859	49,920	13,859	49,920	13,859
IV F-stat	4.655	0.03	6.480	2.758	16.16	8.165	14.53	9.700
Controls					X	X	X	X
Leave-One-Out							X	X

Note: See Table 1.2 for detailed notes, all of which apply here as well. The sample excludes municipalites in the six northern border states (Baja California, Chihuahua, Coahuila, Nuevo León, Sonora, and Tamaulipas).

Table A.5: Excluding modern retail employment from the instrument

Reduced-form regression:								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Products:	All	Energy	Goods	Services	Goods	Services	Goods	Services
Δ (Predicted local labor demand)	-0.00240	0.161***	-0.0505	0.0678**	-0.0980***	0.0592	-0.0865***	0.0876*
	(0.0305)	(0.0508)	(0.0318)	(0.0297)	(0.0280)	(0.0365)	(0.0293)	(0.0480)
IV regression:								
Δ (Log wage bill per worker)	-0.009	0.880*	-0.183*	0.279	-0.188***	0.118	-0.171***	0.169
	(0.116)	(0.512)	(0.0953)	(0.177)	(0.0467)	(0.0862)	(0.0516)	(0.112)
Observations	85,867	3,445	64,373	18,049	64,373	18,049	64,373	18,049
IV F-stat	13.23	3.544	14.71	10.61	43.78	24.87	41.75	27.13
Controls					X	X	X	X
Leave-One-Out							X	X

Note: See Table 1.2 for detailed notes, all of which apply here as well. The shift-share instrument excludes employment in the modern retail sector (supermarkets, mini-supers, department stores).

Table A.6: Price changes immediately following the Peso Crisis

	(1) All products RF	(2) All products IV	(3) Goods RF	(4) Goods IV
Δ (Predicted local labor demand)	0.0234		0.0435	
	(0.0512)		(0.0595)	
Δ (Log wage bill per worker)		0.0190		0.0346
		(0.0393)		(0.0439)
Observations	27,736	27,736	23,734	23,734
IV F-stat		10.61		11.18

Note: Each observation is a barcode-by-store-level product. Price changes are differences in the 3-month averages of Q3 1994 and Q2 1995. Standard errors are clustered at the metropolitan area level. Bartik shocks are constructed using 3-digit industry codes at metropolitan-area-level using Economic Census data for 1993 and 1998. All regressions control for product category fixed effects. Housing is excluded. RF: reduced form regression. IV: instrumental variable regression.

Appendix B

Additional Materials for Chapter 2

B.1 Data Appendix

Cognition data

Participants in KLPS-5 were administered a modified version of the Harmonized Cognitive Assessment Protocol (HCAP), a neuropsychological assessment battery that comprises tests and items in common with other international partner studies of the US Health and Retirement Study (HRS) (**gross˙midlife˙2024**). The standard HCAP cognitive battery was adapted to the Kenyan context and the Swahili language. Confirmatory factor analysis modeling that generates the general and component-specific scores was conducted using Mplus software, version 8.7 (41). We describe the details of the HCAP test and adaptations made below.

- **Swahili Mental State Exam (SMSE):** This test is an adaptation of other mental state exams that is adapted for Swahili-speaking populations (such as those in KLPS), evaluating domains such as temporal orientation (today's date, month, year, and season), spatial orientation (place), registration, attention, recall, and speech.
  - The question "What state are we in right now?" was replaced with "What county are we in right now?" since there are no states in Kenya. Due to recent redistricting in Kenya, it is more likely that participants would know their county rather than the province.
  - Acceptable answers for "Which season is it?" were expanded to include the "long rain season," the "short rain season," and the "January season," rather than just traditional US seasons like winter, spring, summer, and fall (autumn), which are not locally relevant in East Africa.
  - Since most surveys were conducted either outside or in homes, questions like "What floor are we on?" and "Name of building?" were replaced with "What is this place we are in used for?" and "What is the name of this city/place/town?"

- Participants surveyed in the Innovations for Poverty Action (IPA) office were asked “What is the name of the building we are in now?” instead of “What floor are we on?”
- The backwards spelling of “world” was omitted because the Swahili translation for world has multiple syllables. We did not replace this item as we did not find an equivalent Swahili word.
- In the repeat a phrase task, ”No ifs ands or buts” was replaced with ”Tupe tupate tumpatie Taaka” because the English phrase does not have a Swahili equivalent that is a tongue twister.

- **Object Naming:** This test involved three questions on items that should be considered common for the participants.
  - “Coconut” was replaced with “watermelon” because coconut is not common in most rural parts of Kenya, and watermelon is common across all parts of Kenya.
  - “Prime Minister” was replaced with “The President of Kenya.”
- **Symbol Cancellation Test (SCT):** This cognitive assessment tool evaluates visual selective attention, requiring participants to identify and mark specific target symbols from a larger array of distractor symbols within one minute.
- **Community Screening Instrument (CSI-D, 4 items):** This is a test commonly used for detecting dementia in community settings.
  - “Point to elbow” was changed to “Point to neck” because the word for elbow is not widely used in Swahili or local languages and dialects.
  - “Point to the window and then the door/ceiling.” If there was no window, then participants were asked to point to the ceiling and then at the door. If the survey was conducted outdoors, then participants were asked to point to the sky and then to the floor.
- **Raven’s Progressive Matrices (RPM) Test:** This cognitive test measures abstract reasoning and non-verbal intelligence, particularly fluid intelligence, which involves reasoning and problem-solving skills that are independent of acquired knowledge.
- **Digit Span Tests:** These are cognitive assessments used to measure verbal short-term and working memory. The numbers were read to the participants in a monotonic nature (said in a continuous prose without tonal variation).
  - **Forward Digit Span:** Assesses short-term auditory memory by asking participants to repeat a sequence of digits in the same order.
  - **Backward Digit Span:** Assesses working memory and cognitive control by asking participants to repeat a sequence of digits in reverse order.

- **Similarities and Differences: Judgment and Problem Solving:** This test involves identifying similarities and differences between various concepts, objects, or ideas. The test is used to assess abstract reasoning, verbal comprehension, and conceptual thinking.
  - The "roses and daisies" comparison item was changed to "kale and spinach" because these flower names are not common in rural Kenya.
- **10 Word List Recall:**
  - "Butter" was replaced by "milk" since butter is not common in rural Kenya.
- **Found a Lost Child:** This test assesses cognitive abilities in real-world contexts, using cognitive functions like problem solving, executive functioning, and social cognition.
  - The scenario involving "found a lost child" was changed to "found a lost child that is looking for their parents" because children often walk around alone without adults in Kenya, making the original scenario less unusual.
- **Logical Memory Story #2 (East Boston Story):** Part of the Wechsler Memory Scale (WMS), this test assesses verbal memory through story recall.
  - The name and city were replaced with a localized version, "Anne Wanyonyi from East Nairobi." "Cafeteria" was replaced with "a kitchen," "State Street" with "Main Street," and dollars with shillings (the local currency unit).
- **Constructional Praxis:** This test assesses the ability to construct or assemble objects, involving motor skills and visual and spatial perception.
  - The diamond shape was replaced with a triangle since diamonds are not commonly taught in Kenyan schools.
- **Go/No-Go Trial:** This test measures the ability to control impulsive responses, assessing executive functioning, response inhibition, attention, focus, and impulse control.
- **Hand Movement Sequencing Test:** This assessment evaluates executive functions related to the frontal lobe, including testing motor planning and execution, inhibitory control, attentional flexibility, and working memory.
- **Token Test:** This assessment evaluates auditory comprehension and language processing abilities.
  - Diamonds were replaced with triangles (for reasons noted above).

Construction of job task indices

We follow the procedure adopted by the World Bank's Skills Toward Employment and Productivity (STEP) surveys, which is an initiative to generate internationally comparable data on skills and tasks available in developing countries, described in Dicarlo et al. (2016). Specifically, we construct the following three job task indices. Each one of the following questions is standardized into a variable with mean 0 and SD 1. Then the components of each index are added to yield the index, ensuring that all questions receive equal weights.

The analytical Index is the sum of all below:

- Length documents read as part of work. Measured from 1-5. 1= One page or less, 2= 2 to 5 pages, 3= 6 to 10 pages, 4= 11 to 25 pages and 5= More than 25 pages.
- Do tasks that involve more than 30 min thinking as part of work. Measured from 1-5. 1= Never, 2= Less than once a month, 3= Less than once a week but at least once a month, 4= At least once a week but not every day and 5= Every day.
- Do math as part of work. Measured from 1-5. Sum of answer choices of a) Measure or estimate sizes, weights, distances, etc., b) Calculate prices or costs, c) Use or calculate fractions, decimals, or percentages, d) Perform any other multiplication or division and e) Use more advanced math, such as algebra, geometry, trigonometry.
- Learning new things at work. Measured from 1-5. 5= Every day, 4= At least once a week, 3= At least once a month, 2= At least every 2-3 months, 1= Rarely or never. These measures are reverse-coded to be consistent with other measures that contribute to this index.

The interpersonal index is the sum of all below:

- Supervise as part of work. Measured from 0-1. 1= yes.
- Contact with clients as part of work. Measured from 1-10. Specifically the survey asks: "Using any number from 1 to 10, where 1 is little involvement or short routine involvements, and 10 means much of the work involves meeting or interacting for at least 10-15 minutes at a time with a customer, client, student or the public, what number would you use to rate this work?"
- Making a presentation at work. Measured from 0-1. 1= yes.
- Collaborating with co-workers at work. Measured from 1-5. 1= Never, 2= Less than once a month, 3= Less than once a week but at least once a month, 4= At least once a week but not every day and 5= Every day.

The Routine/Manual index is the sum of all below:

- Freedom to decide as part of work (reverse-coded). How much freedom do you (did you) have to decide how to do your work in your own way, rather than following a fixed procedure or a supervisor's instructions? Use any number from 1 to 10, where 1 is no freedom and 10 is complete freedom. These answers are reverse-coded for constructing the index.

- Frequency of repetitive tasks as part of work (reversed). Measured from 1-4. 4= Almost all the time, 3= More than half the time, 2= Less than half the time, 1= Almost never. These answers are reverse-coded for constructing the index.

- Operate heavy machinery as part of work. Measured from 0-1. 1= yes.

- Drive as part of work. Measured from 0-1. 1= yes.

- Repair/maintain electronic equipment as part of work. Measured from 0-1. 1= yes.

- Work is physically demanding. Measured from 1-10 where 1 is not at all physically demanding (such as sitting at a desk answering a telephone) and 10 is extremely physically demanding (such as carrying heavy loads, construction worker, etc). Higher number means more physically demanding.

Economic preferences

The economic preference module elicits individuals' responses across four different domains:

- Risk preferences: Participants were offers a 50% chance of receiving some amount for sure versus a lottery. Following standard approaches, we keep the 50% lottery amount fixed at 900 Shillings, but we ask a series of questions varying the amount of payment for sure to pin down the individual's certainty equivalent.

- Social preferences: Participants were asked whether they would buy a thank-you gift for stranger who spent an hour of their own time to help them (reciprocity), give to a charity helping people in their ancestral home area, (in-group altruism), and give to a charity helping people in Kenya but not in their ancestral home area (out-group altruism)

- Time preferences: Participants were asked to choose between a payment of a certain amount today versus a higher amount in one month. We keep the amount today fixed at 300 Shillings but vary the amount in the future across a series of questions.

- Ambiguity: Participants were asked to choose between sure probabilities versus unsure ones by playing a game where they drew a ball out of a bag without looking. There are two bags with 10 balls each. In bag 1, out of 10 balls there are 4 red balls and 6 yellow balls. In bag 2, there are also 10 balls, but the number of red and yellow balls is unknown. They can choose a bag from which they are to draw the ball. If they choose bag 1, to win 50 shillings they need to draw a red ball. If they choose bag 2, to win 50 shillings they need to decide a color and draw a ball of that color.

We code up risk and time preferences into continuous variables based on the amount that the participant was trading off for. In total, 9 variables are used from the economic preferences module in our analyses.

Big Five personalities

We measure individuals' Big Five personalities with 3 questions each across five domains.

- Extraversion: Tends to be quiet (reverse-coded), act as a leader, is full of energy
- Agreeableness: Compassionate, rude to others (reverse-coded), assumes the best of people
- Conscientiousness: Is disorganized (reverse-coded), has difficulty getting started on tasks (reverse-coded), is reliable
- Neuroticism: Worries a lot, tends to feel depressed, is emotionally stable
- Openness: Is fascinated by art, music, or literature, interested in abstract ideas, comes up with new ideas

All questions are measured on a 5-point scale from "disagree strongly" (1) to "agree strongly" (5). The score on each personality is the sum of its three measures. We also include questions about fatalism—"is it impossible to escape a predetermined fate" and life satisfaction—"how satisfied are you with life". These yield seven personality variables that we use in our analyses.

B.2 Appendix Figures and Tables

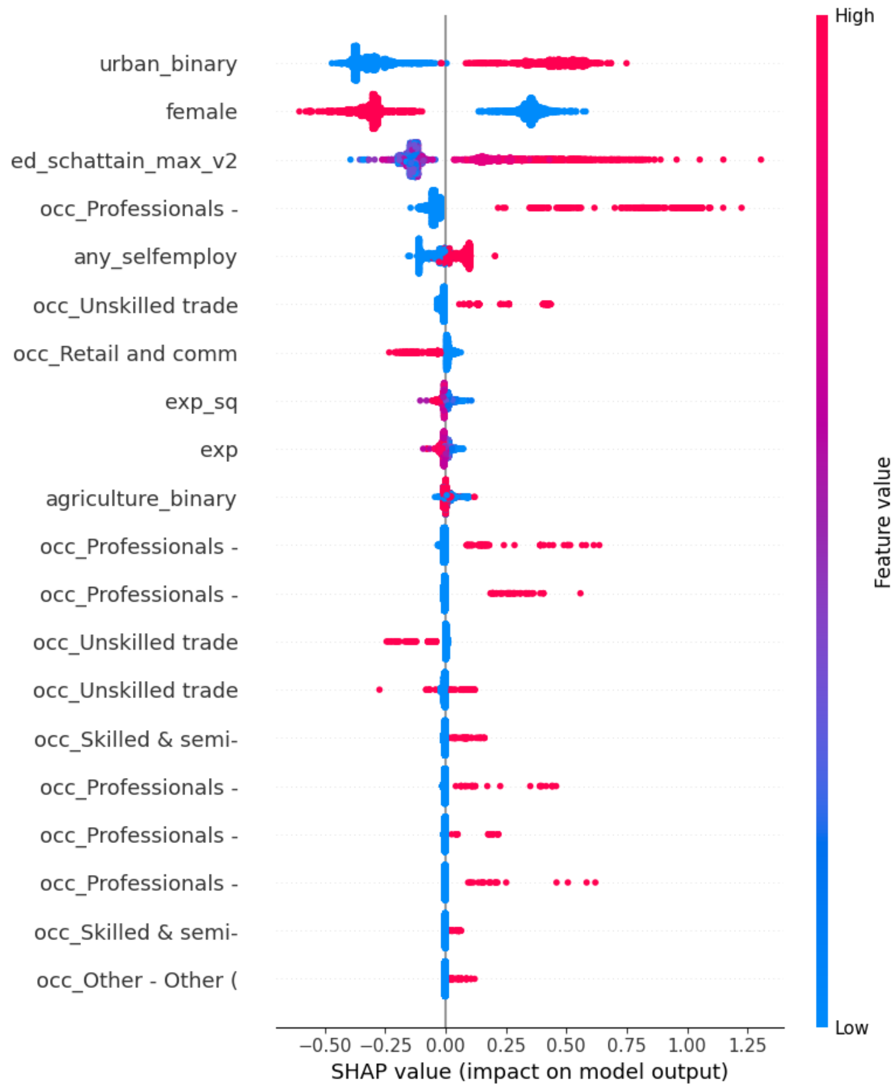


Figure B.1: SHAP summary plot predicting wages with basic set of predictors

Note: See Figure 2.5 notes, all of which apply here. The model includes 55 education, experience, demographics, and occupation variables as predictors of log wages.

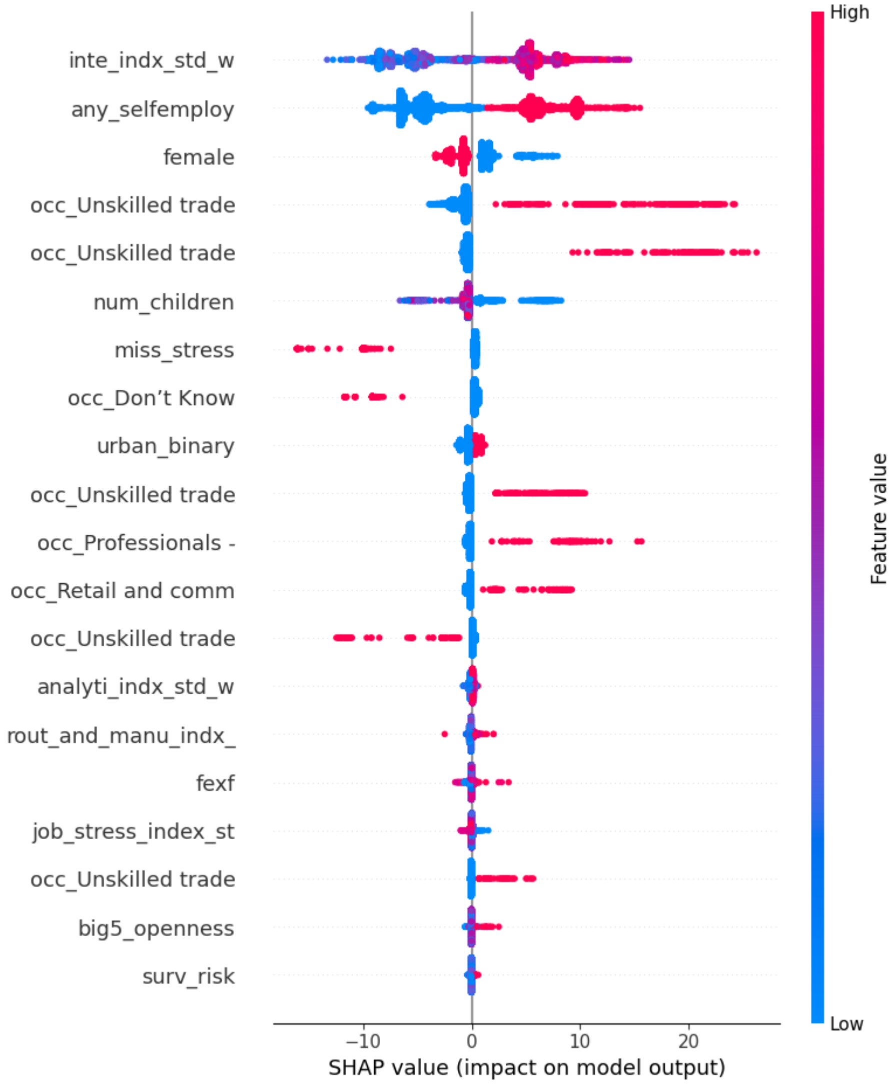


Figure B.2: SHAP summary plot predicting hours worked

Note: Figure shows the impact of key features on predicted hours worked, which are derived from a random forest machine learning model. Features are ranked by their average absolute SHAP (SHapley Additive exPlanations) values, with more important predictors appearing higher on the vertical axis. Each dot represents a single observation, and its position on the horizontal axis reflects the SHAP value (i.e., the contribution of the feature to the prediction for that observation). Dot colors indicate feature values, with higher feature values in pink and lower feature values in blue.

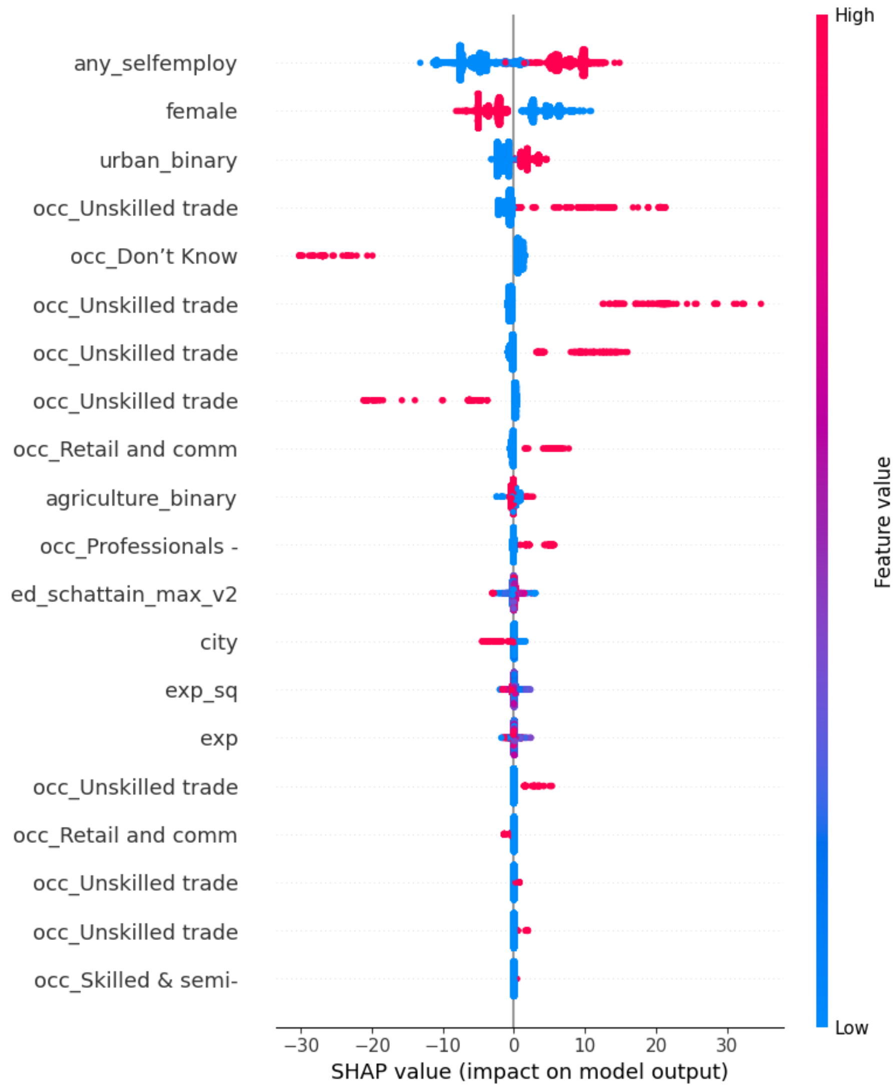
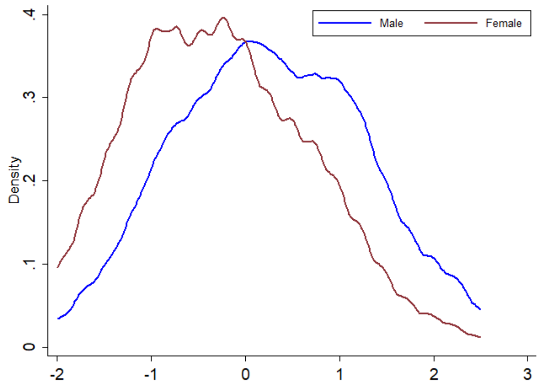
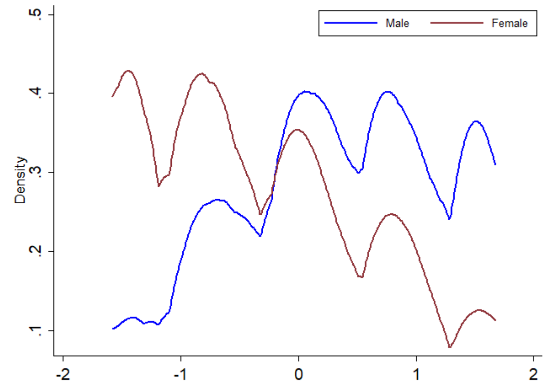


Figure B.3: SHAP summary plot predicting hours worked with basic set of predictors

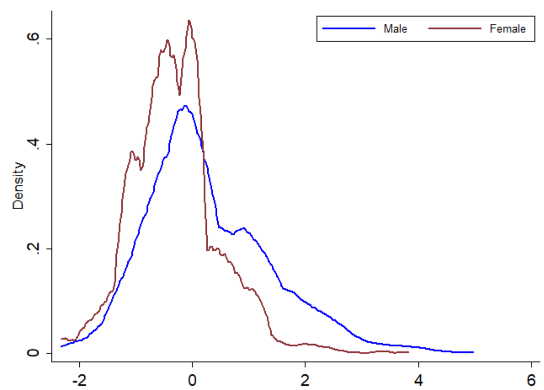
Note: See Figure B.2 notes, all of which apply here. The model includes 55 education, experience, demographics, and occupation variables as predictors of hours worked.



(a) Analytical Index



(b) Interpersonal Index



(c) Routine/Manual Index

Figure B.4: Kernal density plots of task indices by gender

Note: Figures plot the estimated kernal densities by gender for the analytical task index, interpersonal task index, and routine/manual task index.

Table B.1: Differences in job task content by gender

	Mean (female)	Mean (male)	Diff	Std. Error	t
Analytical Index	-0.26	0.26	-0.52	0.03	-20.05
Length documents read as part of work	-0.20	0.19	-0.39	0.03	-15.13
More than 30 min thinking as part of work	-0.16	0.16	-0.32	0.03	-12.01
Do math as part of work	-0.20	0.20	-0.40	0.03	-14.96
Learning new things as part of work	-0.16	0.17	-0.33	0.03	-12.16
Interpersonal index	-0.34	0.33	-0.67	0.03	-26.62
Supervise as part of work	-0.24	0.24	-0.48	0.03	-18.54
Contact with clients as part of work	-0.18	0.15	-0.32	0.03	-12.09
Make presentation at work	-0.21	0.22	-0.43	0.03	-16.39
Collabrating with co-workers at work	-0.33	0.32	-0.65	0.03	-25.04
Routine/Manual index	-0.30	0.27	-0.57	0.03	-22.30
Freedom to decide as part of work	-0.11	0.09	-0.20	0.03	-7.31
Repetitive tasks as part of work	-0.02	-0.04	0.02	0.03	0.57
Operate heavy machinery as part of work	-0.21	0.22	-0.43	0.03	-16.47
Drive as part of work	-0.24	0.25	-0.49	0.03	-19.48
Repair/maintain equipment as part of work	-0.13	0.14	-0.27	0.03	-10.12
Work is physically demanding	-0.11	0.07	-0.18	0.03	-6.62
Observations	5609				

Note: Table shows the gender differences and the t-test statistics for each component of the analytical, interpersonal, and routine/manual task indices.

Table B.2: Groupings of KLPS and Census at 1-digit occupational level

Census 1-digit Grouping	Census (%)	KLPS 1-digit Grouping	KLPS (%)
Agricultural employee/self employed	33.6	Agriculture and fishing	32.5
Profesional/technical/managerial/clerical	24.3	Unskilled trades	19.7
Sales and Services	16.7	Retail and commercial	18.1
Unskilled manual	13.7	Skilled & semi-skilled trades	16.7
Household and domestic	8.2	Professionals	11.2
Skilled manual	2.2	Other occupation	1.8
Other	1.3		

Note: The table presents the distribution of occupations in the Kenyan Census (KDHS) and KLPS data at the 1-digit level.

Table B.3: Oaxaca-Blinder decomposition of wages, using female as the reference group

	Coef.	95% CI Lower	95% CI Upper
With basic controls:			
Difference(Male-Female)	0.780	0.705	0.854
Explained	0.285	0.135	0.435
Unexplained	0.495	0.337	0.652
With full set of controls:			
Difference(Male-Female)	0.780	0.691	0.868
Explained	0.359	0.285	0.433
Unexplained	0.421	0.322	0.519

Note: This table reports the Oaxaxa-Blinder decomposition results of the gender gap in log wages, using female as the reference group. "Basic" controls include eduaction, experience, demographics, and occupation variables. "Full" controls additionally include cognition, personality, economic preferences, job stress, and home care responsibilities.

Table B.4: Oaxaca-Blinder decomposition of hours worked, using female as reference

	Coef.	95% CI Lower	95% CI Upper
With basic controls:			
Difference(Male-Female)	12.099	10.593	13.606
Explained	5.008	1.763	8.253
Unexplained	7.091	3.646	10.537
With full set of controls:			
Difference(Male-Female)	12.099	10.201	13.998
Explained	7.196	5.478	8.914
Unexplained	4.903	2.448	7.358

Note: This table reports the Oaxaxa-Blinder decomposition results of the gender gap in hours worked using female as the reference group. "Basic" controls include eduaction, experience, demographics, and occupation variables. "Full" controls additionally include cognition, personality, economic preferences, job stress, and home care responsibilities.

Table B.5: Gender gap in hours worked: interaction terms

	(1)	(2)	(3)	(4)
Female (intercept)	-13.55***	-46.42**	-4.12	-40.48*
	(4.08)	(22.07)	(4.01)	(22.12)
Interaction terms with Female				
Urban	-3.49	-3.65	-6.14**	-5.09**
	(2.59)	(2.62)	(2.48)	(2.51)
Nairobi	-0.63	-0.20	6.86**	3.20
	(3.09)	(3.05)	(3.22)	(3.02)
Agriculture	3.18	3.11	-2.72	-2.50
	(2.43)	(2.39)	(2.48)	(2.40)
Self-employment/business	3.43*	3.49*	-0.90	0.29
	(1.91)	(1.92)	(1.83)	(1.84)
Educational attainment	0.07	0.62	0.05	1.33**
	(0.31)	(0.58)	(0.30)	(0.58)
General cognition score		8.60		2.33
		(12.66)		(12.10)
Executive functioning		-4.36		-3.11
		(4.96)		(4.95)
Orientation		-1.67		0.37
		(2.35)		(2.37)
Memory		-4.77		-3.13
		(6.86)		(6.57)
Language		-2.95		-2.64
		(3.20)		(3.13)
Visuospatial		-1.07		0.08
		(2.51)		(2.53)
Lagged Raven's		0.33		0.12
		(1.13)		(1.08)
Experience		2.46		1.93
		(1.74)		(1.75)
Experience^2/100		-0.05		-0.03
		(0.04)		(0.04)
Analytical Index		3.15***		3.39***
		(1.14)		(1.16)
Interpersonal Index		-0.33		0.18
		(1.21)		(1.16)
Routine/Manual Index		1.70		2.19**
		(1.20)		(1.02)
N	5772	5772	5772	5772
Model	Regression	Regression	DML	DML
Controls	Basic	Full	Basic	Full

Note: Columns 1 and 3 include the basic set of controls a la Column 4 in Table 2.4, along with the listed interaction terms. Columns 3 and 4 include the basic set of controls a la Column 7 in Table 2.4, along with the listed interaction terms. Columns 3 and 4 run DML regressions following the same setup as in Table 2.5. All standard errors are robust.

Table B.6: Gender gap in wages, winsorized at the 95th percentile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-0.75***	-0.60***	-0.56***	-0.49***	-0.46***	-0.40***	-0.40***
	(0.05)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Educational attainment		0.08***		0.04***	0.02	0.01	0.01
		(0.01)		(0.01)	(0.01)	(0.01)	(0.01)
Experience		-0.13***		-0.04	-0.03	-0.02	-0.03
		(0.04)		(0.04)	(0.04)	(0.04)	(0.04)
$\text{Experience}^2/100$		0.00**		0.00	0.00	0.00	0.00
		(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
Urban		0.58***		0.40***	0.39***	0.34***	0.33***
		(0.06)		(0.06)	(0.06)	(0.05)	(0.05)
Nairobi		0.15**		0.11*	0.12**	0.11*	0.12**
		(0.07)		(0.06)	(0.06)	(0.06)	(0.06)
Agriculture		0.10*		-0.03	-0.04	-0.01	-0.01
		(0.05)		(0.05)	(0.05)	(0.05)	(0.05)
Self-employment/business		0.03		0.26***	0.27***	0.23***	0.19***
		(0.04)		(0.05)	(0.05)	(0.05)	(0.05)
General cognition score					0.18***	0.15***	0.03
					(0.02)	(0.02)	(0.25)
# Older adults in HH					0.03	0.03	0.05
					(0.06)	(0.06)	(0.06)
# Children in HH					-0.04***	-0.04***	-0.04***
					(0.01)	(0.01)	(0.01)
Analytical Index						0.10***	0.08***
						(0.03)	(0.03)
Interpersonal index						0.10***	0.09***
						(0.03)	(0.03)
Routine/Manual index						-0.00	0.02
						(0.02)	(0.02)
Job stress index							-0.13***
							(0.02)
N	5063	5063	5061	5061	5061	5061	5061
r2	0.08	0.22	0.32	0.35	0.36	0.37	0.39
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: Table follows the exact same specification and structure as Table 2.3, but the log hourly wage variable is winsorized at the 95th percentile which is 5.75 (equivalent to 313.43 KSH or 2.24 USD). All standard errors are robust.

Table B.7: Gender gap in hours worked, winsorized at the 95th percentile

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Female	-11.64***	-10.21***	-9.54***	-8.27***	-8.15***	-6.62***	-6.56***
	(0.93)	(0.90)	(1.01)	(0.97)	(0.97)	(1.01)	(1.03)
Educational attainment		0.58**		0.42	0.32	0.28	0.30
		(0.28)		(0.26)	(0.26)	(0.26)	(0.27)
Experience		1.15		1.08	1.10	1.36	1.41*
		(0.95)		(0.85)	(0.85)	(0.85)	(0.83)
Experience^2/100		-0.03		-0.02	-0.02	-0.03	-0.03
		(0.02)		(0.02)	(0.02)	(0.02)	(0.02)
Urban		6.05***		2.14*	1.99	1.02	0.83
		(1.27)		(1.23)	(1.24)	(1.26)	(1.27)
Nairobi		-1.62		-1.64	-1.57	-1.75	-1.66
		(1.56)		(1.41)	(1.41)	(1.38)	(1.37)
Agriculture		0.33		-2.66**	-2.76**	-2.25*	-2.39**
		(1.19)		(1.15)	(1.16)	(1.17)	(1.16)
Self-employment/business		12.23***		10.98***	11.00***	10.20***	10.26***
		(0.88)		(0.99)	(0.99)	(1.00)	(1.01)
General cognition score					0.71	0.28	13.05**
					(0.52)	(0.52)	(5.70)
# Older adults in HH					-0.91	-0.88	-0.81
					(1.18)	(1.18)	(1.19)
# Children in HH					-0.26	-0.29	-0.32
					(0.24)	(0.23)	(0.23)
Analytical Index						1.45**	1.48***
						(0.57)	(0.57)
Interpersonal index						3.06***	3.13***
						(0.61)	(0.61)
Routine/Manual index						0.29	0.24
						(0.54)	(0.53)
Job stress index							0.37
							(0.50)
N	5773	5773	5772	5772	5772	5772	5772
r2	0.04	0.10	0.22	0.26	0.26	0.27	0.28
Occupation_FE			Yes	Yes	Yes	Yes	Yes
Personality							Yes
Preferences							Yes

Note: Table follows the exact same specification and structure as Table 2.4, but the weekly hours worked variable is winsorized at the 95th percentile which is 89 hours. All standard errors are robust.

Table B.8: Gender wage gap, controlling for hours worked

	(1)	(2)	(3)	(4)
Female	-0.59***	-0.48***	-0.65***	-0.51***
	(0.05)	(0.05)	(0.04)	(0.04)
hours_lastweek	-0.01***	-0.01***		
	(0.00)	(0.00)		
N	5061	5061	5061	5061
Model	Regression	Regression	DML	DML
Controls	Basic	Full	Basic	Full

Note: Columns 1 and 2 of this table follows the exact same specification as Columns 4 and 7 of 2.3, but include weekly hours worked as an additional covariate. Columns 3 and 4 are the DML analogues of Columns 1 and 2, again with weekly hours worked as an additional covariate. Because DML allows the relationship between covariates and the outcome to be nonlinear, there is no coefficient on hours_lastweek in Columns 3 and 4. All standard errors are robust.

Appendix C

Additional Materials for Chapter 3

C.1 Data Appendix

This appendix contains additional details of the Lagos Trader Survey (LTS) data. More details can be found in Startz (2024). The data used in this paper comes from surveys conducted in Nigeria as part of the Lagos Trader Project, which was founded in 2014 by Meredith Startz and Shelby Grossman as a research project to understand topics in economics and political science. The sample includes 599 respondents who were already surveyed in previous rounds, as well as a new sample of 946 respondents. Original LTS respondents were surveyed in Round 1 between July and November 2023; new respondents were surveyed in Round 2 between April and August 2024. To obtain the new sample, enumerators conducted a listing exercise of major markets in Lagos to obtain a complete enumeration of all shops in those markets. Specifically, the listing was conducted in Lagos Island, Trade Fair, Computer Village, Alaba, and several others plazas in Lagos Mainland, Lekki, and Iyoki. The listing included any business in permanent physical premises, but excluded mobile vendors and hawkers. It also included shops that were empty, locked, or occupied by offices or service-providers. During this enumeration, research assistants generally did not interact with people in the shops, but rather documented the location and type of activity based on passive visual assessment. When multiple categories of products were being sold, enumerators were instructed to label the shop according to the one that appeared to be the main line of business.

The listing exercise provides us with a list of more than 68,000 shops from which we sample for the survey. Because the listing did not identify individual shops, sampling took the form of designating how many shops selling a given product should be interviewed from each location. For instance, the sampling might dictate that we interview the owner of one of the five shops selling shoes on the first floor of a particular building. To ensure that the selection of shops within that group was not biased, enumerators were given randomized "directions"—to enter the designated floor of the plaza, turn in a randomly selected direction (left or right), and count to the Xth shop (randomly selected out of the total number) selling the designated type of product. There was likely error involved in locating the Xth shop from the designated entrance, but the goal was only to prevent enumerators from defaulting to convenience sampling based on respondents' observable characteristics or location within buildings. Enumerators were instructed to approach sampled shops in person and ask to speak with the owner of the business. In the event that the owner was not available, they were instructed to ask when the owner would be available and return at a later time. On some very rare occasions, enumerators were given permission to interview a manager instead of the owner—this occurred only when the owner was not likely to be at the shop any time within the next few months (e.g., if they lived abroad) and the manager was the person responsible for sourcing stock for the business.

Once the owner was identified, they were provided with an information sheet describing the study, as well as contact information for the PIs and local project manager. In addition to standard informed consent procedures, potential respondents were told that they could contact the PIs or project manager if they had any questions or wanted to verify the legitimacy of the research before agreeing to participate. Many did so, and this appears to have alleviated skepticism about the research and concerns about fraud. A separate team of research assistants audited a 12% random sample of interviews to verify data consistency and ensure that survey protocols were followed. When the respondent's answers differ between the original interview and the audit interview, the research assistant would ask about the reasons for the discrepancy and update the record if the participant confirms that their most recently given answer is correct.