Unpacking Entrepreneurial Ability*

Brenda Samaniego de la Parra†

Ajay Shenoy‡

August 6, 2025

Abstract

The canonical treatment of entrepreneurial ability is as a univariate black box whose specific role in production is assumed positive but not explicitly modeled. We develop a model with multiple stages of production to obtain a microfounded, multidimensional estimator of entrepreneurial ability in single-establishment retailers. A successful retailer must 1) attract customers, 2) source and stock inventory, and 3) choose the right mix of products. Our method estimates ability across all three dimensions. Though each dimension affects profit, they do not aggregate into a scalar measure of ability. We apply our estimator to novel high frequency data collected from small retailers in Lusaka, Zambia. The dimensions of ability are only weakly correlated, suggesting most entrepreneurs excel in one but not all aspects of production. Counterfactual calculations based on a simple univariate estimate of TFP would yield misleading results and, if used to target interventions, could actually lower social welfare.

JEL Codes: D24, L23, L81, C13

*Previously circulated under the title "Measuring and Estimating Retail Productivity" and "Unpacking Total Factor Productivity in Retail"

†University of Notre Dame

‡University of California, Santa Cruz. Corresponding Author. Email at azshenoy@ucsc.edu. Phone: (831) 359-3389. Website: http://people.ucsc.edu/∼azshenoy. Postal Address: Rm. E2455, University of California, M/S Economics Department, 1156 High Street, Santa Cruz CA, 95064. We would like to acknowledge generous financial support for this project from the International Growth Center's Small and Growing Business Fund; Private Enterprise Development in Low Income Counties (PEDL); and the U.C. Santa Cruz Institute for Social Transformation. We gratefully acknowledge the many helpful comments and suggestions we received on this project from David Lagakos, Imran Rasul, Chris Woodruff, Jakob Brounstein, and many participants in seminars and conferences at UCSC, Yale, IFS/UCL/LSE, Edinburgh, Imperial College London, and Cardiff University. We also thank our excellent research assistants Daniel Riegels, Nitya Tripathi, and Emerald Waters; Chisenga Chimenge and the entire team at BEMAC Research Solutions, Ltd.; and Aleta Starosta and Heather Huntington for their advice and support.

1 Introduction

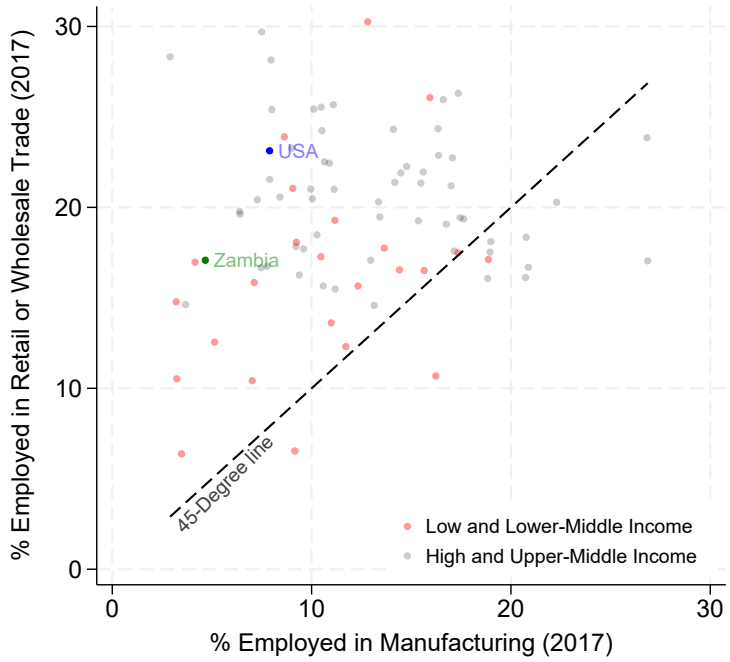
The average firm in a developing country makes far less output per worker than its counterpart in the developed world (Bloom et al., 2010). This fact has motivated a host of interventions targeting small entrepreneurs in developing countries (Quinn and Woodruff, 2019), especially by helping them manage their businesses more efficiently. Yet the results vary dramatically by context (McKenzie and Woodruff, 2017), leaving a puzzle for researchers and policymakers. Part of the puzzle may lie in the absence of a model that explicitly maps the abilities of the entrepreneur to the industry-specific operations of the firm. The standard production framework models the firm as a black box that transforms inputs into output without any notion of the stages in between. By necessity of its generality, it is not microfounded in the actual operations of the firm. In reality, production occurs in multiple, distinct stages, and an entrepreneur's ability to manage each stage may differ. It is far from clear whether an entrepreneur who excels in one stage will excel in the others, or how to identify a specific entrepreneur's weaknesses.

These limitations are especially clear in the retail sector. A retailer does not produce a physical good but a match between a consumer and a good produced by others. The unit of retail output, a sale, does not exist separately from consumer demand. As a result, producing more output requires producing consumer demand, which is distinct from the other half of making a successful match: selecting and sourcing goods produced by others. Each of these distinct stages happens separately and almost certainly requires different abilities. It is crucial to understand each distinct dimension of ability when designing policies to help entrepreneurs in the retail sector, since it would be wasteful to, for example, subsidize the purchases of a shop that is skilled at sourcing and stocking inventory but incompetent at selecting products that customers actually want. In other words, great ability in one stage of production is wasted if not matched in the others.

Getting such policies right would have huge consequences. The retail sector employs more people than manufacturing in nearly every country at every stage of development (see Figure 1).[1] Productivity in the retail sector determines the incomes of the hundreds of millions of workers it employs. It is also crucial for many billions more who rely on the retail sector to source and stock the food and medicine and innumerable other goods they consume. But while in the U.S. over 80 percent of those employed in retail work for large firms with modern systems for managing logistics and inventory, the vast majority of retail workers in developing countries are employed in small and relatively unproductive "traditional" shops Lagakos (2016). Understanding what drives productivity in these small independent retailers is thus critical to improving the average incomes of these economies.

[1]The figure combines retail and wholesale trade because we are unaware of any internationally comparable dataset that measures them separately. Among countries that report separate statistics, wholesale is a much smaller share of employment.

Figure 1: Retail/Wholesale Trade Employs More People than Manufacturing Nearly Everywhere



Note: Data from Inklaar et al. (2023).

In this paper, we explicitly model production in the retail sector as a multi-stage process in which firms' output is the result of their performance in each of several interconnected operations. Identifying and modeling each stage requires an in-depth understanding of the shop's operations. The necessary data to develop such insights was previously unavailable even for developed economies, let along for small and growing businesses in developing countries. We therefore construct a unique panel dataset from a geocoded high-frequency survey of single-establishment shops in Lusaka, Zambia. This survey is one part of an extensive two-year field project that also includes a detailed baseline survey of assets, suppliers, and managerial practices. The sample includes shops at many levels of sophistication, from stores with printed signs and paid employees to outdoor vegetable stands. The panel contains responses to a daily survey answered one week per month for 8 months.

We establish seven stylized facts about the small-scale retail sector. First, there is wide variation in profits, with the top 10 percent earning more than four times as much as the median. Second, there is also variation in the prices shops pay for their inventory, even when we restrict attention to homogeneous goods (e.g., fresh ginger root). Third, shops that stock the same product may choose different varieties (e.g., red versus white onions), and those varieties earn different average profits. Fourth, stockouts are very common. Fifth, different shops make dif-

ferent choices of restocking plans (less frequent larger purchases versus more frequent smaller purchases), and this choice is correlated with the cost of restocking. Sixth, there are large differences in the number of customers that visit each shop, and in the use of even basic means to attract more customers (such as posting printed signs). And finally, nearly all of the variation in shop-level profits can be explained by shop fixed effects (at least within the 8-month window of our panel). Taken together, these facts establish that entrepreneurs differ in how they run three key aspects of their business: which goods to stock and what price is paid for each; the logistics of restocking those goods; and the methods (and effectiveness) with which they market their shop to customers. Moreover, fixed characteristics of the entrepreneur explain their vastly different outcomes.

We then construct a model that captures each of these operations and use the model to estimate entrepreneurial ability in all three dimensions. We assume that a productive retailer must 1) attract customers, 2) source and stock inventory, and 3) choose the right mix of products from the right suppliers. Each is a distinct and sequential stage of production where the outcome, which depends in part on one dimension of entrepreneurial ability, then becomes a state variable in the next stage. First, the shop's location, prices, and advertisement expenditure affect the number of potential customers that visit the store. Holding these inputs fixed, a shop that attracts more customers has higher "marketing ability." Second, the shop's labor, capital, and "logistical ability" determine its capacity to transport and display inventory. Given this capacity, and the expected number of customers and expected sales per customer, it chooses a total stock of inventory and how often to restock within a period (say, one week). Third, holding fixed total inventory, the restocking schedule, and the number of customers, the shop must allocate its inventory across individual goods to minimize both unsold inventory and stock-outs. To succeed, it must accurately forecast which products will be in demand, and source them from the cheapest suppliers. A shop that minimizes the share of inventory unsold has high "inventory choice ability."

Models are meant to simplify complex interactions in the economy with the goal of extracting useful lessons and driving policy. However, the canonical assumption that entrepreneurial ability or productivity is unidimensional is not innocuous. We show that allowing for three dimensions of entrepreneurial ability in a three-stage model of retail leads to novel predictions that run counter to what would be derived from a standard model where entrepreneurial ability is represented by a single TFP-style scalar (as in, for example Lucas Jr, 1978). First and most important, it predicts that each dimension independently raises profit, but their impacts cannot be aggregated into any univariate index. Second, the marginal product of additional inventory may be higher *or lower* for shops with higher inventory choice ability. The intuition is that, for levels of inventory beyond what is needed to satisfy expected sales, each additional dollar of inventory serves as buffer stock. An entrepreneur with greater ability to forecast demand for each good will suffer fewer costly stock-outs, lowering the returns to the buffer. This is in

stark contrast to standard models, where higher ability or productivity always implies a higher marginal product.

Third, the three-stage model predicts that, all else equal, an entrepreneur with higher logistical ability will choose a *lower* stock of capital. The intuition is that if the shop can source more inventory with the same capital and labor, but cannot expect to sell the additional inventory, then the optimal response is to achieve the same inventory with fewer resources. This is the opposite of what the standard model predicts, where the most productive firms choose the most labor and capital. Yet we show that in our dataset, capital is indeed negatively correlated with logistical ability, as would be expected if the measures of ability are not strongly correlated. This result has serious implications for the literature studying the aggregate cost of cross-firm misallocation of capital and other inputs. A researcher could measure large and spurious misallocation because shops with high logistical ability but low inventory choice ability would appear highly productive in the standard model, but would rationally choose lower levels of capital and thus appear constrained. For the same reason, an intervention to provide grants of capital, if targeted to shops measured as more productive based on the single-stage model, could be partly or wholly wasted.

Finally, our model of multi-dimensional ability predicts that an increase in logistical ability alone may actually lower social welfare. To understand why, consider a grocer who, through better logistics, can stock more inventory but who happens to be exceptionally bad at predicting which vegetables are in demand (i.e., low inventory choice ability). Much of the extra inventory may go unsold and be discarded. If there are negative externalities from wasted goods (e.g., increased garbage collection) the net impact may be negative. This again contrasts with the standard model, where increasing a firm's entrepreneurial ability simply increases output. As we explain below, these differences have critical implications for how standard univariate metrics of productivity or ability may be (mis)interpreted, and how policy recommendations based on these metrics may be invalid.

Using the model and the novel panel data from retailers in Lusaka, we estimate the three dimensions of ability. We show that the key conditions for the predictions of our three-stage model to differ from that of the standard, univariate model of prduction hold in our data: the three dimensions of ability are largely independent of one another, and the independent variation captures meaningful information about productivity. These are the necessary conditions for the prediction that shops with higher logistical ability choose less capital, and that is indeed what we find. We also show that estimates of ability constructed from only the first seven rounds of data are better predictors of gross profit (value-added) in the final round than 1) standard measures of managerial practices, and 2) a standard, univariate estimate of TFP.

Taken together, these results imply that relying a standard one-stage model for counterfactual and policy analysis could yield highly misleading results. We demonstrate this point by simulating a hypothetical reallocation of capital from firms with low TFP to those with high

TFP. We compare the estimated gain from a marginal reallocation of capital (1 Zambian kwacha from the bottom to the top of the distribution of MPK) under a standard one-stage model (along the lines of Hsieh and Klenow (2009) or Restuccia and Rogerson (2008)) to the estimated gains in our model. We find that the standard model would overestimate the impact by a factor of 8.

We then compare the impact of a small capital grant targeting firms with high marginal product of capital (MPK) where the targeting relies on the standard one-stage model instead of the multi-stage model. We show in our own sample that the intervention would mistarget nearly one quarter of the sample if based on MPK calculated from a standard model. The aggregate impact would be 11.5 percent larger if the grants were targeted correctly. Finally, we show that a capital grant is the wrong intervention for the vast majority of the sample. A grant that marginally increased fixed advertising investment (e.g., signs advertising the goods and their prices) would have a bigger impact than a marginal increase in physical capital on 84 percent of the sample.

Our paper joins a small but growing literature that explains productivity in developing countries by modeling how information and transaction frictions hinder specific operations of the firm. Startz (2024) applies a structural model to original data on clothing retailers from Lagos to understand how face-to-face travel can mitigate a moral hazard problem in sourcing high-quality products, and to quantify the gains from eliminating moral hazard. Bassi et al. (2023) applies a model of task assignment to novel time-use data from manufacturing firms in Uganda to quantify the potential gains from greater within-firm specialization. Other work studies the role of contracting frictions in driving vertical integration (Boehm and Oberfield, 2020) and relational contracting (Macchiavello and Morjaria, 2021). What distinguishes our work is a focus less on common external frictions than on heterogeneity in how individual entrepreneurs manage the frictions affecting each operation of the firm. In this, our work bears some resemblance to Bernard et al. (2022), who propose a model of production networks where firms differ in their production efficiency and their ability to form relationships with customers and suppliers. Our work differs in both modeling the specific operations (e.g., attracting customers versus finding low-cost suppliers) and deriving firm-level estimates of ability in each dimension.

Our paper also contributes to a subset of the literature on management that shows how good managers alter specific operations within the firm. Adhvaryu et al. (2022) show that managers with high attention and control are able to rapidly learn how to improve the operations of their production line. Adhvaryu et al. (2024) show that managers can weather unexpected absences in their production line by forming relational contracts with other managers to borrow and lend workers. Anderson et al. (2018) show that entrepreneurs in South Africa randomly assigned to receive marketing skills training achieve higher profits through higher sales, while those assigned to financial skills training achieve higher profits through lower costs. Brooks et al. (2018) find that entrepreneurs in Kenya who are assigned local mentors achieve higher profits, mainly through lower costs, with evidence suggesting that mentors told mentees where

to find low-cost suppliers. Our work differs in representing entrepreneurs in a richer, multi-dimensional way, allowing their abilities to to affect firms' operations at different stages in distinct ways, and using a structural model to estimate these different dimensions of managerial ability from differences in realized outcomes.

2 Data and Stylized Facts

2.1 Description of the Lusaka Retail Productivity Study

The biggest challenge to studying retail is the lack of comprehensive data, especially developing countries. Though many surveys of small and micro enterprises contain retailers, they mainly capture general outcomes like sales and wages. Many key aspects of the retail operation, like the number of customers or the replacement value of goods sold, are not asked in any existing survey. Even the questions that are asked often have recall horizons too long to produce reliable data in a context where many entrepreneurs do not keep written records of even their sales and expenses.

The first key contribution of this project is a high frequency panel survey of small single-establishment shops in Lusaka, Zambia. The project collected a wealth of data on baseline characteristics, followed by 8 rounds of high-frequency data on dynamic variables needed to estimate the model. The high-frequency panel surveyed each shop daily for one week per month, aiming to minimize recall error and accurately capture highly variable concepts such as the number of customers or the cost of goods sold.

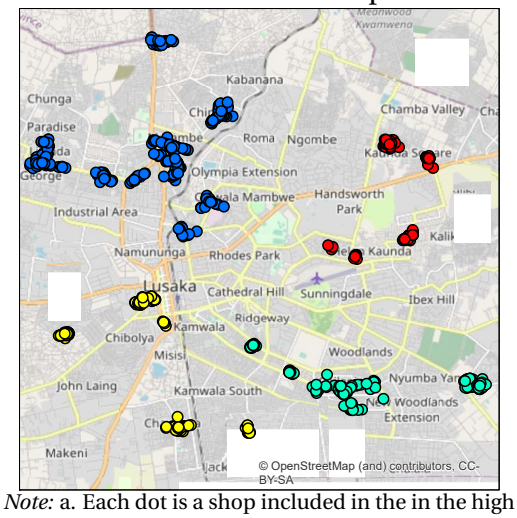
Our study focused on 25 markets in Lusaka, Zambia. We selected these markets as key areas where single-establishment, retail shops are clustered. The study began with a census of eligible retail establishments within the core of each market, yielding roughly 3,000 shops. Within this sample, we restricted attention to two key industries, food and pharmacy, to collect an extensive in-person baseline survey. The baseline survey collected data on the shop's premises, assets, workers, management practices, top products, sales, number of customers, expenses, value of inventory, sources for inventory (names and locations of suppliers and wholesale markets), and restocking. The final sample was roughly 1,000 shops. Figure 2.b shows the locations of the sample shops.

We then ran an eight-month panel phone survey to follow the shops included in the in-person baseline.[2] Each shop was assigned to one of four "clusters" based on their market. Each cluster was surveyed for one week per 28-day period. Each day an enumerator would call each shop to ask roughly 5 minutes of questions about the day's business: hours open, labor hours, sales revenue, cost of goods sold, number of customers, and cost of any goods purchased for

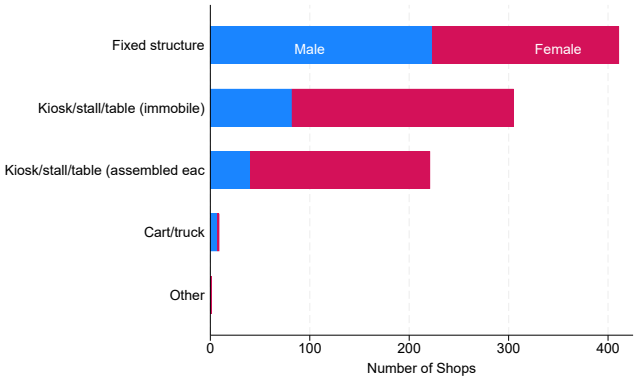
[2] After an initial round where we either surveyed shops in-person or asked them to fill out self-surveys, we moved to a phone format for the remaining 7 months.

Figure 2: Location and Composition of the Sample

a. Locations of Shops



b. Sample Counts by Structure and Gender



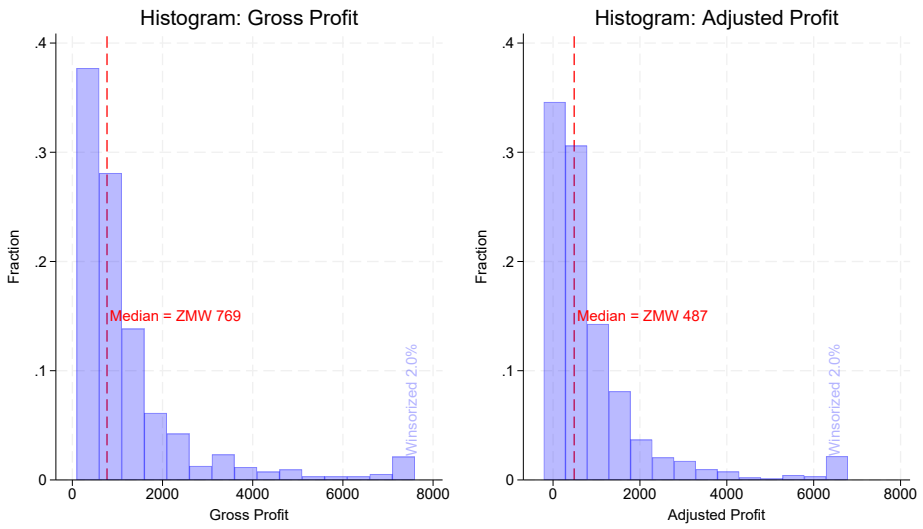
Note: a. Each dot is a shop included in the in the high frequency panel. b. Structure types are fixed (a building with walls), a kiosk that is fixed in place (usually made of metal), a kiosk or table that is taken home by the entrepreneur, and a cart or truck or other vehicle.

resale. Some questions only needed to be asked on specific days, like the level of inventory at the beginning and end of the week. Questions that needed to be asked only once per week—for example, the prices and costs of top products, or whether any equipment had been bought or sold—were spread out during the week to minimize the length of any one survey. If a respondent could not be contacted on one or more days, they would be asked about the days missed during the next successful survey. Respondents who could not be contacted at all—for example, those who never picked up the phone—would be visited in-person at the end of the week to collect the full week's data. This mixed-method approach balanced the need to minimize the lag between the response period and the date of survey while keeping as many shops as possible within the sample. The final sample contains roughly 950 shops that answered at least one survey, though the number that consistently answered across all days in any one round is generally 700 to 800.[3]

Figure 2.a shows the diversity of the types of establishments. Although all shops are independent single-establishment shops selling food, cosmetics, or pharmaceuticals, they take many forms. Though the plurality have a fixed structure, more than half are kiosks or stalls (either immobile or assembled daily). Most shops in the sample are run by a woman, though men run most of the more established shops with fixed structures.

[3]We have far fewer responses in the first round, as the self-survey format was generally unsuccessful. The number of observations in Round 1 is 526.

Figure 3: There is Wide Variation in the Profitability of Shops



Note: The histograms plot two different definitions of profit over one week. Gross profit is defined as sales minus the cost of the goods sold. Adjusted profit is gross profit minus depreciated inventory and the fixed cost of restocking (delivery, transport, and time). Both variables are Winsorized at the top and bottom 2 percent. All numbers are in current Zambian Kwacha (ZMW).

2.2 Stylized Facts of the Small-Scale Retail Sector

Using this data we establish several stylized facts about the outcomes and practices in the micro-retail sector in Zambia.

Fact 1 *There is wide variation in the profitability of shops*

Every shop in our sample is a single-establishment retailer that prior literature would classify within the "traditional" sector (Lagakos, 2016). Yet there is enormous variation in their profits. We define two measures of profit:

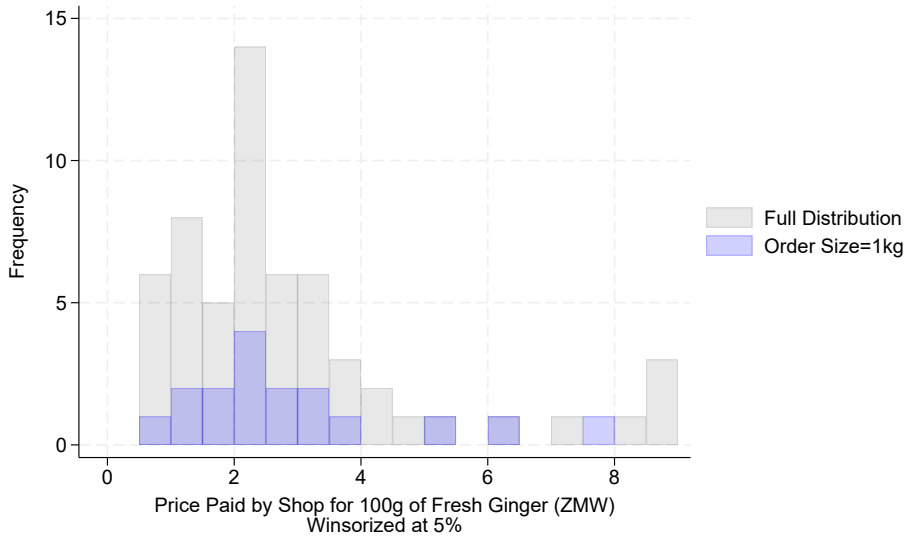
1. Gross profit: sales over one week minus the cost of the goods sold
2. Adjusted gross profit: gross profit minus the value of goods lost to theft or spoilage over the week, and minus the cost of delivery, transport, and time spent restocking during the week.[4]

We average both across all 8 rounds and Winsorize at the 2 percent level. Figure 3 shows that there is wide variation in the sample. Though half of shops earn less than 500 Zambian Kwacha (ZMW) per week in adjusted profit, the top 25 percent earn twice as much, and the top 10 percent earn more than four times as much.

Fact 2 *Some shops can source the same goods at far lower cost.*

[4]We value each hour spent restocking at the wage that the respondent estimates they would have to pay to hire a casual worker. If the respondent was unable to answer the question, we imputed the average response among all respondents in the market.

Figure 4: Prices Paid for Inventory Vary Drastically Even for Highly Homogenized Products



Note: Price data were collected from mid-January through early March during the in-person baseline survey conducted.

Part of the reason for the differences in profits may be differences in the prices paid by the shops for the goods that they resell (the "order price"). Differences in order prices appear in even the narrowest and most mundane niches of Lusaka's retail sector. For example, Figure 4 shows there is wide variation in the price paid for fresh ginger. Though the modal price is roughly 2 Zambian kwacha (ZMW) per 100 grams, some shops pay half that price while others pay two or even four times as much. The differences cannot be explained by the size of the order. They remain even when the distribution is restricted to orders of 1 kilogram.

Fact 3 *Sellers choose different goods, and some choices appear more profitable*

Figure 5 shows that even among shops that sell onions there are differences in the varieties they choose. Nearly all stock white onions, while roughly half stock red onions.[5] Roughly 47 percent stock only one variety. Although both varieties have comparable sales, red onions are more profitable. Though these descriptive statistics do not prove that switching from white to red would raise profits, the patterns are consistent with that interpretation.

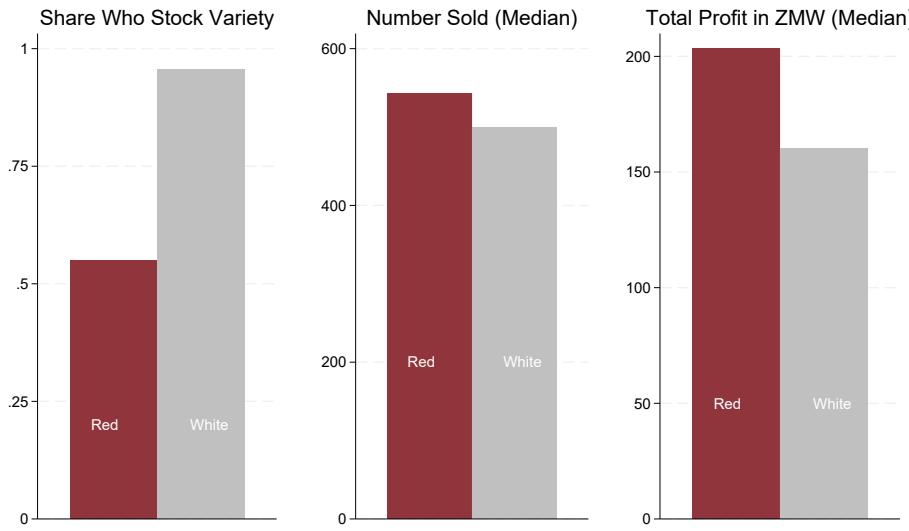
Fact 4 *Stockouts are common*

Across all shops and rounds of the monthly survey, roughly 68 percent report a stockout on at least one day over the 7 day response window, and 7.5 percent report a stockout every day.

Fact 5 *Shops make different choices about the size and frequency of each restocking, and the choice seems correlated with the fixed cost of restocking.*

[5]During the survey, respondents were given the option of choosing between white, yellow, and red onions. We subsequently discovered that white onions (as described in the U.S.) are not sold in these markets. Sellers describe as "white onions" what in the U.S. would be called a yellow onion. We have combined white and yellow onions under the category of white onions because the difference in classification is purely a consequence of enumerator choice, not actual differences.

Figure 5: Sellers Choose Different Goods (Variety of Onions) and Reap Different Outcomes



Note: Sellers who report selling onions were asked about all varieties that they stock. All data were collected from mid-January through early March during the in-person baseline survey conducted. "Number Sold" and "Total Profit" are conditional on stocking the variety, and refer to sales and profits from that variety.

For each shop-round we calculate 1) the number of days that the shop reported buying goods for resale ("restockings"), 2) the average value of the goods purchased at each restocking, and 3) the average fixed cost of delivery, transport, and time per restocking.[6] The left-hand panel of Figure 6 shows that there is wide variation in how frequently shops choose to restock. Roughly equal shares (about 20 percent) will restock once, twice, or thrice per week. Roughly one-quarter of the sample will restock at least five times per week. Shops that restock more frequently will on average buy less per restocking, suggesting there is a choice between frequent small purchases versus infrequent large purchases. The right-hand figure suggests that, after controlling for the log of sales, a larger fixed cost of restocking is correlated with larger purchases and fewer restockings (though the latter correlation only appears for very high fixed costs). The pattern is consistent with the idea that shops that must pay a higher fixed restocking cost will choose a larger, less frequent restocking plan.

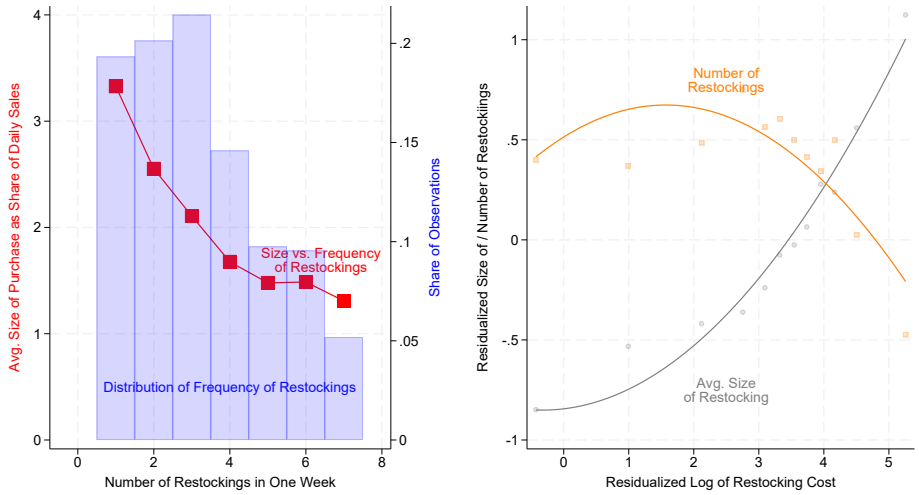
Fact 6 *There is wide variation in the number of customers visiting each shop, and this variation is correlated with observable choices like having a printed sign.*

Figure 7 shows there is enormous variation across shops in the average number of customers per day.[7] The modal shop has just over 30 customers per day, but the bottom decile has less than 10 customers per day and the top decile nearly 100. Even a truly basic form of advertising—having printed or professional-looking signs on the exterior of the shop—predicts a higher number of customers (a difference of roughly 10 for the median shop).

[6] See Appendix C for the precise definition.

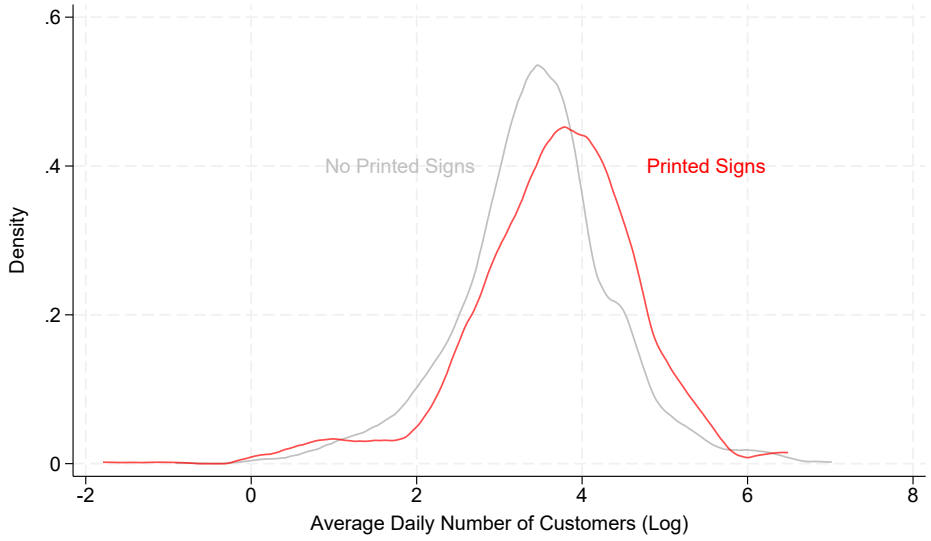
[7] A customer is defined as someone who enters a shop (if it is enclosed) or approaches (if it is a table or kiosk) regardless of whether they buy anything.

Figure 6: Shops Choose Different Restocking Plans and Face Different Costs Per Restocking



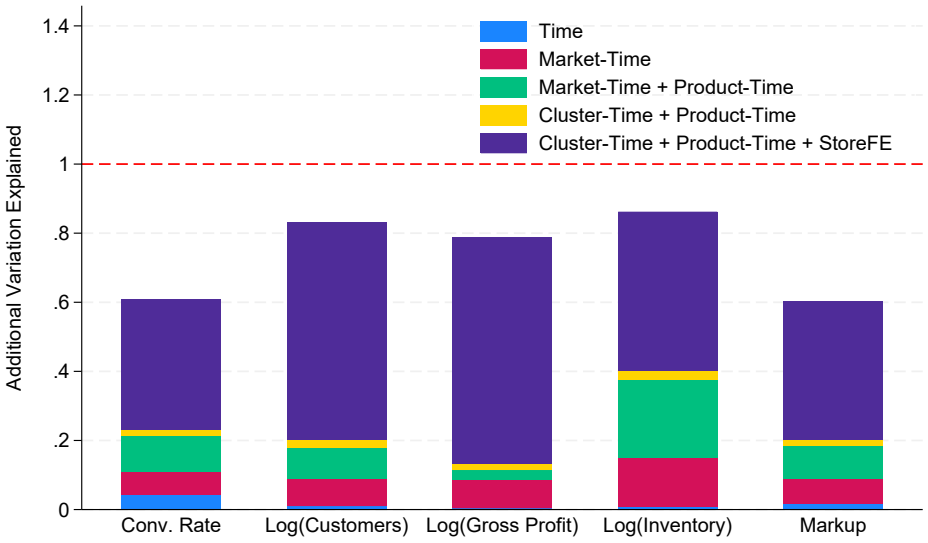
Note: A "restocking" is defined as any day when the retailer reports buying goods for resale (e.g. 7 implies they restock every day). The "size of purchase" is the value of goods bought during the average restocking (as a share of average daily sales). The restocking cost captures the fixed cost of the average restocking (financial cost of delivery or transportation, time cost of picking up goods and physically restocking the shelves). Both panels restrict to shops that restock at least once during the reference week of the shop-round. The right-hand panel residualizes all variables after controlling for the log of total sales during the week. The left-hand panel calculates the average value within each discrete value of the frequency of restocking, while the right-hand panel bins the residualized outcomes by quantiles of the residualized cost of restocking.

Figure 7: There is Wide Variation in Daily Customers, but Shops with Printed Signs Attract More



Note: Each day the shop was asked how many "customers" there were, defined as any individual who entered (for a fixed establishment) or approached and looked at the goods (for a table or kiosk) regardless of whether they bought anything.

Figure 8: The Bulk of the Variation is Explained by Store Fixed-Effects

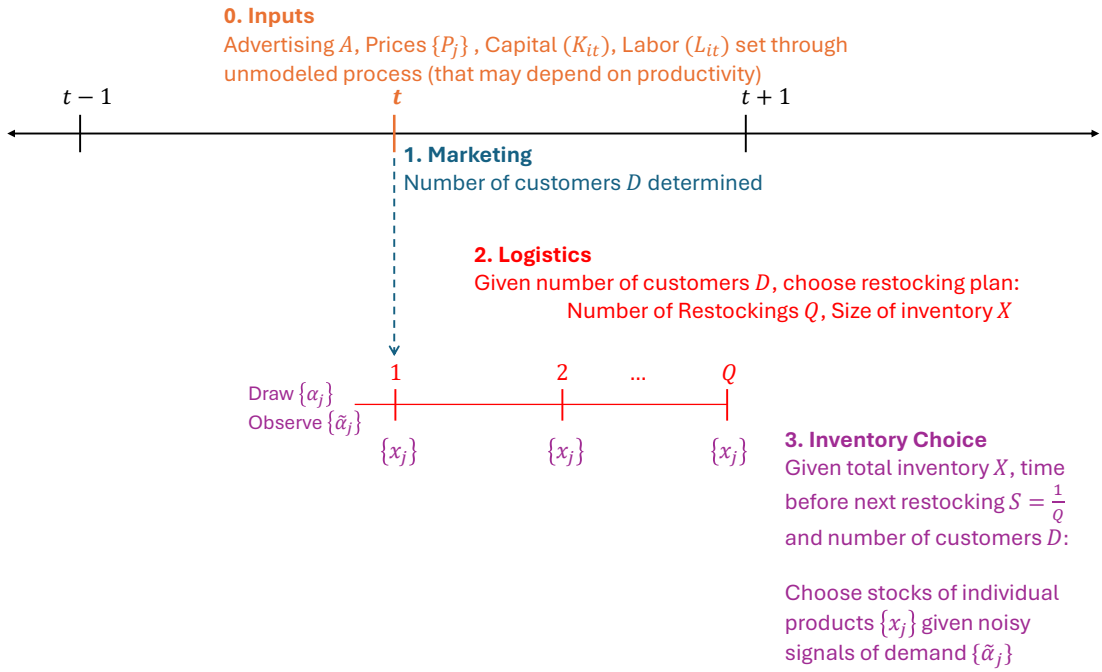


Note: For each outcome we estimate a sequence of regressions that incrementally add fixed effects. The first regression contains only time fixed-effects, the next market-time fixed effects, the third market-time as well as product-time fixed effects, and so on. For each outcome we graph the R-squared of the regression with just time fixed effects, then stack the change in R-squared as we add an additional layer of fixed-effects. "Cluster" refers to a geographical unit constructed by k-cluster analysis on the GPS coordinates of the shop. Clusters were constructed so that all shops within a cluster lie no more than 750 meters from one another.

Fact 7 *The variation in outcomes is largely explained by fixed characteristics of the shop or entrepreneur.*

These patterns suggest that there is wide variation in how (and how well) different shops are managed. And it is indeed the identity of the shop, more than any feature of the market or the product line, that determines the shop's outcomes. Figure 8 shows the incremental variation explained in a series of key outcomes as we add additional fixed effects. The first specification contains only time fixed-effects, the next market-time fixed effects, the third market-time as well as product-time fixed effects, and so on. Each layer of each bar shows the additional variation explained by that specification. The full specification, which includes store fixed-effects, explains most of the variation for all outcomes. In all cases, the final specification (which includes store-fixed effects) explains more than twice as much variation as the next-best specification. Though these fixed-effects are not the same as entrepreneurial ability—they could include differences in slow-moving inputs like capital and labor—they are independent of the location or industry of the shop. It is something about the shop, rather than its location or industry, that predicts its success. That suggests entrepreneurial ability may play a major role in explaining which shops succeed or fail.

Figure 9: Model Timing



3 Model

We build a model to broadly capture the retailers' operations based on these stylized facts. In the model, a successful retail operation is one where the entrepreneur is able to mitigate several key frictions: the information friction of what goods are in demand and where to buy them cheaply; the transaction friction of paying a fixed cost to restock; and the information and travel frictions that arise because customers must learn about and travel to the shops that stock the goods they want at the best price. An entrepreneur who, all else equal, mitigates a specific set of frictions can be said to have high ability in that operation. Throughout this section, we use "shop," "retailer," and "entrepreneur" interchangeably, as this is a model of single-establishment retailers.

The design of any model is a balance of trade-offs, and ours prioritizes making the minimal assumptions needed to derive three intuitive equations that can be estimated using fixed-effects regressions. This balance reflects its purpose: it is a platform for defining and estimating each dimension of entrepreneurial ability, not a comprehensive theory of retail strategy. There are many key choices made by a retailer—how they set prices and select a location, for example—that we assume merely satisfy certain criteria but otherwise are not modeled. Since these choices are observed, they can be controlled for in the estimation.[8] These unmodeled

[8]This is analogous to how the production function literature does not model the choices of capital and labor

features are determined in a "Stage 0" that precedes all else.

The model then proceeds in three parts. The first part models the arrival of customers at the shop, which depends on its location, prices, marketing expenditures, and "marketing ability." Given the number of customers, the second part models the shop's choice of restocking schedule—frequency and size of inventory—as constrained by its capacity to source and stock inventory, which depends on capital, labor, and "logistical ability." The total number of customers, restocking frequency, and total inventory are in turn inputs into the third part, which models how the firm chooses which products to stock. The shop observes a noisy signal of the demand for each product, and must choose stocks to avoid both stock-outs and unsold inventory. High-ability entrepreneurs can more accurately forecast demand for individual goods, and source them from low-cost suppliers. Total sales depend on total inventory, the number of customers, prices and markups, and "inventory choice ability."

The timing is as shown in Figure 9. In each period:

0. Shop location, prices, costs, advertising expenditures, capital, and labor are determined through some unmodeled process. These quantities may be correlated with any or all dimensions of ability.

1. The firm makes all choices related to marketing (prices and advertising) which determine the number of customers. The firm is then committed to these choices for some period of time, and the number of customers is known with certainty.

2. The firm chooses a restocking schedule and overall inventory size subject to its logistical capacity (taking as given the number of customers)

3. The shop draws a noisy signal of demand for each product. At each restocking, it chooses the quantity of each good to stock given its signal, the overall size of inventory, the number of customers, and the time until the next restocking. If the firm stocks out of any good, it loses all sales for that good until it can restock again.

We solve the model backwards starting with the selection of inventory.

3.1 Inventory Choice

There are J goods that the shop could potentially stock. At each restocking the shop chooses a stock x_j for each good j, paying the cost κ_j and charging the price P_j for each unit. Customers arrive continuously at a rate D, which is fixed within the period t, known with certainty, and predetermined by the choices made in the marketing part of the model (see below). Each customer buys the same quantity α_j/J, which is fixed within the period but unknown at the time the

beyond imposing the minimal restrictions on those choices necessary to nonparametrically control for them. See the review in Ackerberg et al. (2007) for examples.

shop makes its choices. The J in the denominator reflects that as the space of possible goods increases, each individual good accounts for a smaller share of consumer demand. The version of the model presented here assumes prices do not affect the intensive margin of how much each customer buys, only the extensive margin of how many customers enter the store (see Section 3.3). But Appendix B.1 shows that allowing for an intensive margin is straightforward and does not change the final estimating equation.

At the beginning of each period t the shop draws a new demand, α_j, and observes a signal

$$\tilde{\alpha}_j = \alpha_j + \sigma u_j \tag{1}$$

where u_j has a standard normal distribution that is iid across goods and time. The term σ, which is assumed to be common across goods but may vary across shops, determines the signal-to-noise ratio. Either through experience or effort, a shop with a smaller σ can make a more precise forecast of demand for each good. For simplicity, we assume the shop has an uninformative prior, and that conditional on the signal, prices are uninformative about the true value of α_j.[9] We abstract from learning and assume that the shop does not use realized sales to inform its beliefs at subsequent restockings within the period.[10] Then the signal is the only information available to the shop. Its posterior belief is $\alpha_j \sim N(\tilde{\alpha}_j, \sigma^2)$ for all j.

The shop knows with certainty that it will restock again after an interval of length S and may select inventory equaling a value X, both of which are pre-determined by the solution to the logistical problem (see below). The total demand for good j between the current choice and the next restocking is $DS\alpha_j$. If the shop ever runs out of a good (a stock-out) it loses the sales it would have earned from subsequent demand for the good. It loses these sales until the next restocking. Finally, assume for now that any inventory that is unsold at the end of the period is lost. We show in the next section that, under appropriate assumptions, the solution to this restricted one-period optimization coincides with that of the multi-period optimization.

Under this setup, the quantities in the model are dynamic, in that the stock of good j is changing within each period and equal at time s to $x_{j,s} = x_j - Ds\alpha_j$. But this dynamic process actually collapses to a static and thus tractable decision problem. The shop maximizes expected sales:

$$\max_{\{x_j\}} \mathbb{E}_{\{\alpha_j\}} \left[\sum_{j=1}^{J} P_j \min \left\{ x_j, DS\alpha_j / J \right\} \right] \tag{2}$$

subject to the spending constraint

$$\sum_{j=1}^{J} \kappa_j x_j \; = X \tag{3}$$

[9] Allowing for normally distributed priors would not change anything substantive as long as the priors are common across goods.

[10] Alternatively, one can assume the shop draws a new α_j and $\tilde{\alpha}_j$ at every restocking.

In the main text we assume that each shop has a single fixed markup

$$P_j = \mu\kappa_j$$

that may vary across shops but is the same across goods. In Appendix B.4 we show how this assumption may be relaxed given additional data.

Let Φ and ϕ be the distribution and density functions of a standard normal random variable, and let λ be the Lagrange multiplier on the adding-up constraint (3). Then the optimal choice for any good j satisfies the first-order condition

$$\lambda = \mu\left[1 - \Phi\left(\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J}\right)\right] \tag{4}$$

$$\Rightarrow x_j = \frac{\tilde{\alpha}_j DS}{J} + \frac{\sigma DS}{J}\Phi^{-1}\left(1 - \frac{\lambda}{\mu}\right) \tag{5}$$

Define the expected total per-customer demand as

$$\zeta = \frac{1}{J}\sum_{j=1}^{J}\kappa_j\tilde{\alpha}_j \tag{6}$$

and the unweighted average cost of goods as

$$\dot{\kappa} = \frac{1}{J}\sum_{j=1}^{J}\kappa_j$$

Subbing (5) into the budget constraint (3) yields

$$\lambda = \mu\left[1 - \Phi\left(\frac{X - \zeta DS}{\sigma DS\dot{\kappa}}\right)\right] \tag{7}$$

This expression shows that one common intuition from the one-stage model—that holding fixed inputs, a more productive firm has higher marginal product—does not apply in this model. A shop with a lower σ will have higher marginal product *if and only if* its overall per-customer spending on inventory is less than expected per-customer demand. Shops with precise signals that are spending too little overall can make use of the marginal dollar more efficiently. But if the shop is already spending too much (more than needed to satisfy overall demand), the dollar is more likely to be wasted because the shop is not stocking out frequently or by large amounts. A shop with imprecise signals, on the other hand, is probably stocking out of many goods even when its overall spending exceeds overall expected demand. The extra dollar will let it hold larger buffer stocks to compensate for its relatively poor choices.

Equations 7 and 4 imply that

$$\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J} = \frac{X - \zeta DS}{\sigma DS\dot{\kappa}} \stackrel{!}{=} \upsilon \tag{8}$$

which we define as the aggregate surplus stock υ. The expression implies that the normalized difference between expected demand for good j and actual stock is the same across goods, and equal to the cost-adjusted normalized difference between total expected demand and total inventory spending.

By substituting Equation 8 into the expression for sales, we show (in Appendix A.1) that

Proposition 1

$$\mathbb{E}[Y] = \mu\Big\{X - \sigma\dot{\kappa}DS\left[\Phi\left(\upsilon\right)\upsilon + \phi\left(\upsilon\right)\right]\Big\} \tag{9}$$

When J is large,

$$Y \approx \mathbb{E}[Y]$$

The intuition for this last result is that although the demand for any individual good is uncertain, overall sales is known with certainty. The manager of the shop knows how noisy her signals are and thus knows exactly what share of her overall inventory will be sold. Put another way, she knows half her inventory will go unsold—just not which half. In later sections we will maintain this large-J assumption to keep the model analytically tractable.

Equation 9 shows that, holding all else fixed, shops with higher costs $\dot{\kappa}$ and noisier signals (higher σ) have lower sales. We define inventory choice ability as

$$\mathcal{I} = \frac{1}{\sigma\dot{\kappa}}$$

which implies that the value of inventory left unsold just prior to restocking again is

$$X - \frac{Y}{\mu} = \frac{DS}{\mathcal{I}}\left[\Phi\left(\upsilon\right)\upsilon + \phi\left(\upsilon\right)\right] \tag{10}$$

which is inversely related to inventory choice ability. This equation in its current form cannot be estimated because υ is unobserved, but we show in the next section how υ can be replaced with a nonparametric function of terms that are observed.

3.2 Logistics

3.2.1 Logistical Capacity

The shop must source, store, and stock its goods efficiently. Its capacity to buy and maintain inventory depends on

- the vehicles used to find and transport inventory, how many shelves it installs for display, and the equipment used to help put goods on the shelves; and
- how many workers it hires to pick up goods and stock the shelves.

Suppose the various forms of capital are $\{K_a\}$ where, for example, K_1 is the quantity of vehicle-hours, K_2 the quantity of shelving, and so on. Let $\{L_b\}$ likewise be the various forms of labor, with L_1 the hours spent picking up goods, L_2 the hours spent stocking shelves, and so on. Suppose that the store's capacity to move inventory to source and stock inventory is

$$C = \left[\sum_a \beta_a^{\frac{1}{\epsilon}} K_a^{\frac{\epsilon-1}{\epsilon}} + \sum_b \nu_b^{\frac{1}{\epsilon}} L_b^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \tag{11}$$

Let p_a^K and p_b^L be the prices of these varieties of capital and labor, and suppose the shop spends $K = \sum_a p_a^K K_a$ and $L = \sum_b p_b^L L_b$ in total on capital and labor.

A shop with an efficient logistical scheme can source and stock the same quantity of goods using fewer resources. It might, for example, plan its restocking to minimize the number of trips or choose a shelving scheme that maximizes the available space. Suppose that the shop can source and stock 1 dollar of inventory at cost $P_C = 1/Z$, where Z is logistical ability. Then the total quantity of new inventory that the shop can move must be less than or equal to

$$\frac{C}{P_C} \tag{12}$$

Assume that the shop is free to reallocate its spending on labor and capital efficiently. Then it is easy to show that the constraint on new inventory becomes

$$Z\left[\beta K^{\frac{\epsilon-1}{\epsilon}} + \nu L^{\frac{\epsilon-1}{\epsilon}}\right]^{\frac{\epsilon}{\epsilon-1}} \tag{13}$$

We assume for simplicity that $\epsilon = 1$, implying (13) becomes

$$ZK^{\beta}L^{\nu} \tag{14}$$

though it is easy to generalize the derivation and estimation.

3.2.2 Restocking Plan

Taking its capacity as given, the shop will set a restocking schedule and a target level of inventory. To simplify the exposition in the main text we make some critical assumptions:

1. At each restocking the shop can sell its surviving inventory and re-optimize the stocks of individual products (that is, inventory purchases are reversible)

2. The unsold inventory it sells back retains a depreciation-adjusted share of its value $\mathcal{D}(S)$, where $\mathcal{D}'(S)$ is everywhere continuous and negative. This function is identical for all goods.

3. When the shop restocks it draws a new set of product-level demands $\{\alpha_j\}$ and signals $\{\tilde{\alpha}_j\}$, but the arrival rate of customers D is the same

The first two assumptions imply that only the choice of overall inventory and the restocking schedule is dynamic. We show in Appendix A.2 that the stocks of individual goods will coincide with the one-period optimization solved for in the prior section. The third assumption implies the dynamic problem is homogeneous across each restocking. The choice of restocking schedule is equivalent to simply choosing how frequently to restock; and the choice of how much inventory to hold is the same across all restocking events.

Suppose the shop is making these choices to maximize its gross profits over a planning horizon of length 1. Then the frequency of restocking Q is simply $1/S$, the inverse of the time between each restocking. Suppose that each restocking has a fixed cost f that reflects the financial and time cost of transporting the goods. The shop solves

$$\max_{Q,X} Q\left\{Y - \left[X - \mathcal{D}(1/Q)(X - Y/\mu)\right] - f\right\}$$

Subject to:

$$X \leq \mathcal{D}(1/Q)(X - Y/\mu) + HZK^{\beta}L^{\nu} \tag{15}$$

The expression $HZK^{\beta}L^{\nu}$ is simply logistical capacity, as derived in the previous section, and capacity utilization or "hours," which the shop may freely scale down (e.g. by closing the shop on weekends) to save on variable costs. The other term on the right-hand side of (15) is the total inventory purchased at the previous restocking that is neither sold nor lost. The constraint imposes that the level of inventory available after restocking is no greater than the level of inventory left from the previous restocking plus the utilized capacity to buy and stock new inventory. Since the problem is stationary, the new level chosen will be the same as the level previously chosen, and it suffices to choose that one target level of inventory X knowing that sales Y is a function of that level.

In the main text of the paper we make two simplifying assumptions:

1. $\mathcal{D}(1/Q) = 1 - \delta$

2. $H = 1$

These assumptions simplify the exposition and proofs but, as we show in Appendix B, yield nearly identical estimating equations.

The following proposition is critical for the estimation of logistical ability:

Proposition 2 *Assume $f > 0$ and $Q > 0$.*

Then the constraint (15) will always bind:

$$X = (1-\delta)(X - Y/\mu) + ZK^{\beta}L^{\nu}$$

The proof, which follows from the Kuhn-Tucker conditions, is found in Appendix A.3.

Proposition 2 implies that

$$Y/\mu + \delta(X - Y/\mu) = ZK^{\beta}L^{\nu} \tag{16}$$

The left-hand side of this equation is total sales (valued at cost) plus the value of inventory depreciated (unsold inventory discarded or stolen). Aside from the extra term on the left-hand side, this equation resembles the production function in a standard single-stage model. But as we explain below, the difference is crucial in how we interpret the impact of an all-else-equal increase in logistic ability. In Appendix B.2 we show that when depreciation varies by the time between restocking, Equation 16 is nearly identical except that the right-hand side is multiplied by the hours that the shop is open.

The first-order conditions of the problem also imply the following proposition, which is crucial in the estimation of inventory choice ability:

Proposition 3 *Let $m = \mu - 1$, the net markup, and let B denote the average quantity purchased during a typical restocking. Then*

$$\upsilon = \Phi^{-1}\left(\frac{mB - f}{(m+\delta)B - (1-\delta)f}\right) \tag{17}$$

We show in Section 5.3 how Equation 17 can be combined with Equation 10 to derive an estimating equation.

3.3 Marketing and Customer Acquisition

We adapt the model of Eaton and Kortum (2002) by allowing customers living at a set of origins indexed by o to travel to a set of shops (indexed by n).

Each shop has an "attractiveness" R_n that is perceived by each potential customer with noise ξ drawn from a Fréchet distribution. Potential customers cannot perfectly observe the average price $\bar{P}_n$ charged by any shop. They have a common prior belief, which is lognormal with mean $\log \hat{P}_n$ and variance χ_0^2. At the beginning of the period, consumers receive a common signal

$$\tilde{P}_n = \bar{P}_n e^{\nu_n} \tag{18}$$

where ν_n has a lognormal distribution with a mean of zero and a variance of χ^2. Define the

Kalman gain

$$G = \frac{\chi^{-2}}{\chi^{-2} + \chi_0^{-2}}$$

The posterior belief has a lognormal distribution with mean $G \log \tilde{P}_n + (1 - G) \log \hat{P} n$ and variance $(\chi^{-2} + \chi_0^{-2})^{-1}$.

Customers weigh their posterior belief about the price, adjusted for iceberg travel cost $\tau_{on} \geq 1$, against the perceived attractiveness of their goods. The customer chooses the shop with the highest perceived attractiveness per expected dollar of expenditure, meaning the shop with the highest value of

$$\frac{R_n \xi}{\tilde{P}_n^G \hat{P}_n^{1-G} \tau_{on}} e^{\frac{1}{2}(\chi^{-2} + \chi_0^{-2})^{-1}} \tag{19}$$

Suppose that the draws of ξ are iid with Fréchet parameter ω, and that origin o has a population of potential customers equal to μ_o. By the properties of the Fréchet distribution, the number of customers arriving at shop n from origin o is

$$\frac{\left(\frac{R_n}{\tilde{P}_n^G \hat{P}_n^{1-G} \tau_{on}}\right)^\omega}{\sum^b \left(\frac{R_b}{\tilde{P}_b^G \hat{P}_b^{1-G} \tau_{ob}}\right)^\omega} \mu_o \tag{20}$$

The total number of customers arriving at shop n is the sum of (20) across origins:

$$D_n = \sum_o \frac{\left(\frac{R_n}{\tilde{P}_n^G \hat{P}_n^{1-G} \tau_{on}}\right)^\omega}{\sum^b \left(\frac{R_b}{\tilde{P}_b^G \hat{P}_b^{1-G} \tau_{ob}}\right)^\omega} \mu_o \tag{21}$$

$$= \left(\frac{R_n}{\tilde{P}_n^G \hat{P}_n^{1-G}}\right)^\omega \sum_o \frac{\left(\frac{1}{\tau_{on}}\right)^\omega}{\sum^b \left(\frac{R_b}{\tilde{P}_b^G \hat{P}_b^{1-G} \tau_{ob}}\right)^\omega} \mu_o \tag{22}$$

$$= \left(\frac{R_n}{\tilde{P}_n^G \hat{P}_n^{1-G}}\right)^\omega \Gamma_n \tag{23}$$

where Γ_n is the population-weighted relative distance of shop n from its potential customers. A shop with a more desirable location will have a higher Γ_n.

The baseline estimator models attractiveness as a period-by-period set of positive impressions arising from the shop's marketing ability and through advertising investment. This could be an investment of money (buying ads, paying for signs) or time (distributing fliers). We assume that advertising is a combination of persistent fixed investments (e.g. signs bought or built at baseline that remain throughout the period of observation) and variable monthly ex-

penditures (advertising or replacement of posted prices).

The impact of this investment is scaled by the entrepreneur's fixed marketing ability $\bar{r}_n$, which reflects her skill (choosing more attractive signs, knowing the best time to distribute fliers) and her personality (making potential customers feel valued). Let t be a time subscript, and denote the log of a variable using lower case letters (e.g. $\log R_{nt} = r_{nt}$). Let

$$r_{nt} = \bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + e_{nt} \tag{24}$$

where $\bar{a}_n$ and $\tilde{a}_{nt}$ are the logs of fixed and variable advertising investment, $\bar{r}_n$ is marketing ability, and e_{nt} is a white noise innovation.

We also consider an extension that models attractiveness as a dynamic form of "relationship capital" built up by the firm. Much like the "knowledge capital" proposed by Ehrlich and Fisher (1982), the shop accumulates (or loses) relationship capital through advertising and ability (or their absence). If not maintained, relationship capital depreciates at a rate $(1 - \rho)$. The law of motion for relationship capital is

$$r_{nt} = \bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + \rho r_{n,t-1} + e_{nt} \tag{25}$$

4 Differences with the Standard One-Stage Model

A standard model assumes production happens in a single stage and output is $Y = Zf(K, L)$. Among small firms in a developing country, the Hicks-neutral TFP term may most credibly be interpreted as a measure of managerial ability (as in Lucas Jr, 1978). The predictions from that model have become the basis for policies to improve firm-level outcomes. This section shows that our multi-stage model can make different or even opposite predictions *if the three dimensions of ability are not well correlated*, which is precisely what we show in the empirical results.

4.1 Each Dimension of Ability Independently Raises Output

The first implication of our model is that gross production (sales) is not the best measure of retail production because it neglects that a successful retailer is able to achieve the same sales with fewer costly restockings and less wasted inventory. The better measure is adjusted gross profit, which we define as gross profit (sales minus the cost of goods sold), minus the replacement value of goods lost at the end of each restocking period, minus the fixed cost of restocking. These terms represent profit per restocking, which must be scaled by the number of restockings

over some fixed period (say, one week).

$$\mathcal{P} = Q\left[Y - \frac{Y}{\mu} - \delta\left(X - \frac{Y}{\mu}\right) - f\right] \tag{26}$$

Focus on the case $\delta = 1$ where all goods are lost at the end of the restocking (the general case, after tedious algebra, yields the same conclusions). Equation 26 becomes

$$\mathcal{P} = Q\left[Y - X - f\right] \tag{27}$$

Define the net markup (markup net of average restocking cost) as

$$\mathcal{M}(Z) = \mu - 1 - \frac{f}{ZK^{\beta}L^{\nu}} \tag{28}$$

Assume $\mathcal{M}(Z) > 0$. It is straightforward to show that Equation 27 can be rewritten as

$$\mathcal{P} = D\left\{\mathcal{M}(Z)\zeta - \frac{\mu}{\mathcal{I}}\phi\left[\Phi^{-1}\left(\frac{\mathcal{M}(Z)}{\mu}\right)\right]\right\} \tag{29}$$

$$= \dot{R}e^{\omega\bar{r}_n}\left\{\mathcal{M}(Z)\zeta - \frac{\mu}{\mathcal{I}}\phi\left[\Phi^{-1}\left(\frac{\mathcal{M}(Z)}{\mu}\right)\right]\right\} \tag{30}$$

where

$$\dot{R} = \left(\frac{\exp(\delta_1\bar{a}_n + \delta_2\tilde{a}_{nt} + \rho r_{n,t-1} + e_{nt})}{\tilde{P}_n^G\hat{P}_n^{1-G}}\right)^{\omega}\Gamma_n$$

Equation 30 shows the contribution of each form of ability to adjusted profit. The expression in braces is profit per customer, which is scaled by the number of customers (and marketing ability). An increase in marketing ability will proportionally increase profit by the other term, profit-per-customer, which is a function of the other forms of ability. An increase in inventory choice ability reduces the negative term inside the braces, which represents the value of goods unsold. That is, a firm with higher inventory choice ability earns higher profit because it throws away less of its stock. Finally, higher logistical ability raises $\mathcal{M}$ by allowing the retailer to buy more inventory per restocking, spreading the fixed cost across more goods and (in practice) reducing the number of times the fixed cost needs to be paid.[11] But though all three measures of ability raise adjusted gross profit, they cannot be represented as a scalar-valued function that is independent of inputs, making it clear that our model is not isomorphic to the standard model.

[11] One can show that

$$\frac{\partial\mathcal{P}}{\partial Z} = D\mathcal{M}'(Z)\left\{\zeta + \mathcal{I}^{-1}\Phi^{-1}\left(\frac{\mathcal{M}(Z)}{\mu}\right)\right\}$$

which is positive because we have assumed $\mathcal{M}(Z) > 0$.

4.2 The Marginal Product is Not Necessarily Higher for Shops with Higher Ability

If we take TFP as the standard model's representation of all advantages of the firm (including managerial ability), then the standard prediction is that the marginal product of an input—say, capital—is strictly increasing in TFP:

$$\frac{\partial^2 Y}{\partial Z \partial K} = f_K(K, L) \tag{31}$$

which is strictly positive under the standard assumption that the production function always has positive returns in all inputs. Given two firms with equal levels of capital and labor, a policy that hands out extra capital or labor should be targeted at the firm with higher TFP. In the absence of any policy, an efficient market would likewise allocate more inputs to that firm.

That is not necessarily what is implied by our model of retail. Equation 7 gives the marginal product of inventory. It implies that

$$\frac{\partial^2 Y}{\partial \sigma \partial X} = \phi\left(\frac{X - \zeta DS}{\sigma DS\dot{\kappa}}\right)\frac{X - \zeta DS}{\sigma^2 DS\dot{\kappa}} \tag{32}$$

This term is positive if and only if $X > \zeta DS$. An increase in inventory choice ability—that is, a *decrease* in σ or $\dot{\kappa}$—raises the marginal product of inventory if and only if $X < \zeta DS$. In other words, it is only efficient to give the high-ability shop additional inventory if both shops have less inventory than needed to satisfy expected demand. Giving a shop inventory in excess of expected demand is building its buffer stocks of goods, making a stockout less likely and, if it happens, less costly. The buffer stock is more valuable to a shop that is not very good at predicting which products are in demand. In the limit, a shop with perfect foresight has no need at all for buffer stocks. Conversely, if the level of inventory is less than expected demand, both shops expect to stock out of many goods. The shop that makes better predictions can use the marginal dollar of inventory spending more efficiently because it can predict and counter the costliest stockouts.

4.3 A Shop with Higher Ability May Choose Less Capital

Our model has an even starker contrast with the standard model in the case of capital and labor. Equation 46 implies that in the standard model a firm with higher TFP will (all else equal) choose a higher level of capital because each penny of investment will yield higher return. That result does not hold in our model of retail, as described in the proposition below:

Proposition 4 *Under the assumptions of Proposition 2, an all-else-equal increase in logistical ability Z lowers the shop's marginal product of capital. As a consequence, the firm will decrease its optimal choice of capital K.*

The proposition is proven in Appendix A.5. The intuition is straightforward. Holding fixed the other dimensions of ability (and thus expected demand and the probability of actually selling a dollar of inventory purchased), a firm with higher Z can buy enough goods to satisfy expected demand with less capital. One can use a parallel argument to prove a similar result for labor.

4.4 Raising Only One Dimension of Ability May Reduce Social Welfare

Equation 16 may on first glance seem a small modification to the usual $Y = ZK^{\beta}L^{\nu}$ production function, but its implications are completely different. To see why, let $\mathcal{S} = \delta(X - Y/\mu)$ be the unsold inventory that is spoiled or lost to theft. Then a small increase in logistical ability implies

$$K^{\beta}L^{\nu}dZ = \frac{\partial Y}{\partial Z}dZ + \frac{\partial \mathcal{S}}{\partial Z}dZ \tag{33}$$

An increase in logistical ability will be portioned between increasing actual sales and increasing the quantity of goods spoiled or lost. Put another way, one "dollar" of extra ability will be divided between greater sales and greater waste. The size of the portion going to sales depends on the number of customers (which depends on marketing ability) and the extent to which demand for individual products is matched to the actual stocks (which depends on inventory choice ability). A similar argument shows that the same is true for an all-else-equal increase in capital or labor.

The policy implication is that an intervention aimed at raising Z_{nt} or distributing capital may or may not actually increase sales, and may actually reduce social welfare by increasing the aggregate waste in the economy. This is in stark contrast to the standard model, where higher TFP always raises output and sales and is always welfare-enhancing. The contrast highlights the danger of ignoring the multi-stage nature of production. Making a firm more successful in one stage without improving its success in the others would at best attenuate the impact, and at worst cause more harm than good.

5 Estimation

Assume there is a panel that records observations of each shop $n = 1, \ldots, N$ during each month $t = 1, \ldots, T$. We show how to use the results of Section 3 to estimate each of the three dimensions of ability. We explain how each quantity in the estimating equations is measured in Appendix C.

5.1 Marketing Ability

Since we neither observe nor require origin-destination flows of customers, we control for location using fixed-effects. Let g be a location (a business center, for example) and assume that

the set of potential customers is similar across all shops in g (i.e., Γ_n is similar for all $n \in g$). Suppose that customers get a signal of the average price of each shop during each period, but their prior is fixed (say, by the shop's average prices in the past). Assume the noise in the signal ν_{nt} is white noise that is independent of all current and past variables. Let ϕ_{gt} be a location-time fixed-effect. Denote logs of variables by lower case variables. Then taking the log of (23) and using ϕ_{gt} to control for Γ_n yields

$$d_{nt} = \omega r_{nt} - \omega \left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{34}$$

$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + e_{nt}\Big) - \omega \left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{35}$$

The signal $\tilde{p}_{nt}$ is not observed but can be replaced by its definition in (18).

$$d_{nt} = \omega r_{nt} - \omega \left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{36}$$

$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + e_{nt}\Big) - \omega \left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{37}$$

$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + (1-G)\hat{p}_n\Big) - \omega \left[G\bar{p}_{nt} + \delta_2 \tilde{a}_{nt}\right] + \phi_{gt} + (G\nu_{nt} + e_{nt}) \tag{38}$$

Equation 38 is the structural form of the regression equation. It can be estimated in two steps. The first step is a reduced form estimation that combines the first parenthetical expression into a single shop fixed effect:

$$d_{nt} = \pi_{0,n} + \bar{\phi}_{gt} + \pi_2 \bar{p}_{nt} + \pi_4 \tilde{a}_{nt} + \varepsilon_{nt} \tag{39}$$

The second step is to isolate marketing ability $\bar{r}_n$ by estimating the expression in the first parenthetical expression in Equation 38 with the combined fixed-effect $\hat{\pi}_{0,n}$ on the left-hand side:

$$\hat{\pi}_{0,n} = \pi_0' + \pi_1' \bar{a}_n + \pi_2' \hat{p}_n + \xi_n \tag{40}$$

The residual from this regression

$$\hat{\bar{r}}_n = \hat{\pi}_0 - \hat{\pi}_1' \bar{a}_n - \hat{\pi}_2' \hat{p}_n \tag{41}$$

is some combination of marketing ability and estimation error from the first stage, and thus an unbiased estimator for marketing ability. In practice this expression may be augmented with location or product fixed effects (e.g. when using the instrument approach described below).

The final challenge to estimating Equations 39 and 40 is that the "average" price $\bar{P}_{nt}$ is very likely measured with error. The first source of error is the challenge of measuring units across multiple goods consistently across different shops. The second is the challenge of defining the

average most relevant for consumer demand given that different shops stock different products, and it is typically only possible to measure prices for a few goods (in the case of our survey, the top three by revenue). We address this challenge by instrumenting the average price with the raw price of the top good (unadjusted for units) while controlling for product fixed-effects and the weight of the unit. In some specifications we instrument with the cost (price paid by the shop to its suppliers) to disentangle the impact of prices on demand from the response of prices to changes in demand.

Before applying the two-step procedure to estimate marketing ability, we estimate different versions of Equation 38 excluding the fixed effect. Though its absence will potentially bias the estimates, they are nevertheless informative about the size of G, the strength of the signal that customers receive about prices. If G is large it implies customers are extremely well informed about each shop's current prices and will react to price increases by immediately gravitating towards shops with lower prices. If G is small it implies customers are generally uninformed about prices and only gradually learn which shops have low prices.

Column 1 of Table 1 shows the most basic model. We estimate the partial correlation between the (log) number of customers and the price index after controlling for advertising expenditure and product-round and market-round fixed effects. We instrument for the price index using the price of the top product (controlling for the size of the unit) as described at the end of Section 5.1. The coefficient is negative, as expected, but noisy and statistically insignificant.

To understand why, in the next three specifications we add variables that might proxy for customers' prior expectation of the price. The most basic is the one-month lag of the price index (instrumented with the lag of the price of the good, controlling for the lag of the unit size). Column 2 shows that the lagged price is statistically significant (though still relatively small in magnitude), and after controlling for it the coefficient on the current price index shrinks. That suggests customers are reacting to price increases with a lag, supporting the idea that customers need time to learn about them. To the extend that the lag of prices is correlated with the prior belief about prices, it suggests G is small. Column 3 pushes the idea further by replacing the lag of prices with the full history of prices (for each round t, the average price across all prior rounds $t' < t$) and adjusting the instrument accordingly. The coefficient on the history of prices is even larger than on the one-period lag. The coefficient on current prices shrinks further to almost zero.

Column 4 switches the proxy to the full within-shop cross-time average of prices (using the cross-time average price of the main product as the instrument). This average price is larger and more statistically significant than any other proxy, suggesting that it is the persistent component of prices rather than changes in the price that drives customer behavior. The coefficient on the current price, which actually flips signs to become positive and significant, is consistent with this interpretation. If customers base their choice entirely on prior beliefs, which are

Table 1: Marketing Ability, Part 1: What Drives Customer to Choose One Shop Over Another? (Equations 38 39)

	Fixed-Effects IV						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Price Index	-0.159	-0.038	0.000	0.170**	0.102*	-0.041	0.120
	(0.098)	(0.098)	(0.115)	(0.073)	(0.056)	(0.199)	(0.289)
Advertising Expenditure	-0.014	-0.002	-0.007	-0.014	-0.005	-0.003	-0.014
	(0.022)	(0.024)	(0.024)	(0.022)	(0.010)	(0.010)	(0.019)
Price Index[t-1]		-0.193**					0.074
		(0.080)					(0.162)
Hist. Avg. Price Index			-0.300**				
			(0.133)				
Avg. Price Index				-0.474***			
				(0.151)			
Customers[t-1]							0.234***
							(0.057)
Shops	873	785	805	873	794	792	719
Shop-Rounds	4661	3584	4010	4661	4575	4565	2799
First-Stage F	204.6	62.0	72.2	141.0	230.5	21.7	1.0
Product-Round FEs	X	X	X	X	X	X	X
Market-Round FEs	X	X	X	X	X	X	X
Store FEs					X	X	X
Instrument	Price	Price	Price	Price	Price	Cost	Price

Note: Columns 1—4 show estimates of different versions of Equation 38 excluding the fixed effect. Columns 5 and 6 show estimates of Equation 39. Column 7 shows estimates from the relationship capital model using the estimator in Appendix B.3. All versions instrument for the average price using either the price or the cost of the shop's top product (by revenue) after controlling for the weight or volume of the unit of sale and product-round fixed effects.

based on persistent differences in average prices, then temporary price increases are probably a response by the shop to increased customer demand (rather than the reverse). Column 5 shows that adding a store fixed-effect, which absorbs the average price (and any other persistent features of the store), leaves this positive coefficient largely unchanged. But the coefficient becomes small and statistically insignificant when (in Column 6) we instrument for the price indices with costs, which are less likely to rise in response to customer demand. Finally, in Column 7 we report the estimates of the relationship capital model from the forward orthogonal deviations estimator. The estimates are similar to those in Column 5, but far noisier.

Taken together these estimates imply that beliefs about the relative prices of shops are persistent and will shift only gradually in response to higher prices. Customers know which shop typically has the lowest price but not whether it currently does. This is likely a feature of the context, where shops have no means to broadcast that they have undercut their competitors. Physical stores may post signs advertising the products they sell, but not the current price. And although stalls will often post signs with prices, a customer must approach to read them. There is no widely employed means for advertising prices to a large swath of potential customers.

Finally, all columns show that our measure of monthly advertising expenditure (the inverse hyperbolic sign of all spending on signs, fliers, and other advertising) has no relationship with the number of customers. Though it is possible that the spending is wasted, it is more likely

Table 2: Marketing Ability, Part 2: Isolating Marketing Ability from Average Prices and Fixed Advertising Investments (Equation 40)

	Stage 2 Estimation			Correlation Across Versions		
	(1) Version 1	(2) Version 2	(3) Version 3	(4) Version 1	(5) Version 1	(6) Version 2
Avg. Price Index	-0.415***	-0.449**	-0.445***			
	(0.148)	(0.195)	(0.134)			
Value of Signs	0.050***	0.053***	0.045***			
	(0.017)	(0.017)	(0.013)			
Version 2				0.974***		
				(0.004)		
Version 3					1.150***	1.169***
					(0.023)	(0.024)
Shops	785	783	776	783	767	767
R-Squared	-0.01	-0.05	0.02	0.99	0.87	0.87
First-Stage F	118.84	55.69	111.46			
Product FEs	X	X	X			
Market FEs	X	X	X			
Instrument	Price	Cost	Price			

Note: Columns 1—3 report estimates of Equation 40. Versions 1 uses the fixed-effect estimated from Column 6 of Table 1 as the dependent variable, and Version 2 uses the fixed-effect from Column 7. Version 3 is based on the forward deviations fixed effect (see Appendix B.3. Columns 4—6 show pairwise relationship between the estimates of marketing ability arising from these three versions (based on simple OLS regressions).

that these monthly expenditures reflect the cost of maintenance or replacement (for example, replacing a sign that was damaged between surveys, or replacing replacing placards that display prices). Though this is new advertising expenditure, it is not necessarily new advertising. The original investment in signs, which are not replaced and thus remain fixed across rounds, may be important in attracting customers (as we show in second-stage estimation described next).

Columns 5—7 each estimate the first-stage equation (39). Although our preferred specification is Column 5, we extract the fixed effect from all three to verify that the final measure of marketing ability is robust to different assumptions about the first stage. Columns 1—3 of Table 2 reports estimates of the second-stage estimation (Equation 40). In all cases we instrument for the within-shop cross-round average of the price index using the analogous average of the price or cost of the main product controlling for the unit of sale and good fixed effect, as well as a market fixed effect.

As expected, the coefficient on the price index is large, negative, and statistically significant. The coefficient on our measure of fixed advertising investment (the inverse hyperbolic sign of the value of all signs made or purchased by the shop) is positive and significant. Both estimates are consistent with the hypothesis that beliefs about prices and the most valuable advertising investments are fixed, though we cannot rule out they are simply correlated with some unobserved factor. As long as the unobserved factor is not well correlated with marketing ability, the residual of these regressions is an unbiased estimate of it. The last three columns of the table show the correlation of this estimate across all three versions of the estimation. Versions

1 and 2, both based on Equation 39 but using different instruments for prices, are highly correlated (the r-squared is nearly 1). Their correlation with Version 3, which is based on the forward deviations estimator, is weaker but still high (the r-squared is 0.87).

5.2 Logistical Ability

Equation 16 cannot be estimated directly because Y/μ and $\delta(X - Y/\mu)$ represent sales and depreciated inventory *between restocking events*. Neither quantity is observed, and even if observed would be incomparable across shops with different restocking frequencies. We can normalize these quantities to a fixed period—say, a week—by multiplying both sides of Equation 16 by the number of restocking events Q_{nt} made by shop n in week t. Take logs of both sides of the transformed equation to yield

$$\log \tilde{Y}_{nt} = \log Z_n + \beta \log K_{nt} + \nu \log L_{nt} + \log Q_{nt} \tag{42}$$

where

$$\tilde{Y}_{nt} = Q_{nt} Y_{nt}/\mu_{nt} + Q_{nt}\delta_{nt}(X_{nt} - Y_{nt}/\mu_{nt})$$

Under the assumption that logistical ability Z_n is fixed within shop over the 8 months of our sample, this equation can be estimated using fixed-effects. To account for firms that restock zero times during the week, we replace $\log Q_{nt}$ with a full set of dummy variables $\{\mathcal{Q}_{nt,p}\}$ for all p possible values of Q_{nt}. Since capital is notoriously hard to measure in any context, we follow Collard-Wexler and De Loecker (2016) by instrumenting the level of capital with net purchases of new capital. In some specifications we also control for the (log) hours that the shop is open during the week, which is necessary in the more general model of Appendix B.2 where the depreciation rate increases with the time between restocking events.

Then logistical ability can be estimated as

$$\log \hat{Z}_n = \frac{1}{T}\sum_{t=1}^{T}\left[\log \tilde{Y}_{nt} - \hat{\beta}\log K_{nt} - \hat{\nu}\log L_{nt} - \sum_p \hat{\mathcal{Q}}_{nt,p}\right]$$

Table 3 reports estimates of Equation 42. The outcome is the repurchase price of goods sold over the course of the week plus the value of goods lost to spoilage, theft, or given away. The specification in Column 1 estimates the regression using OLS. Since capital is likely measured with error, we follow Collard-Wexler and De Loecker (2016) and instrument capital with investment. Column 2 instruments with only positive investment while Columns 3 and 4 construct an instrument that measures both positive and negative investment. Columns 1—3 control only for round fixed-effects, while Column 4 controls for round-industry fixed effects (where

industry is the broad category of the shop's best selling product).[12] Column 5 estimates the enhanced model that allows depreciation to vary with the time between restocking (see Appendix B).

The capital coefficient is implausibly small in Column 1. But it becomes large (roughly 0.47) after instrumenting, consistent with a downward bias in OLS caused by measurement error. The labor coefficient is consistently roughly 0.41 in Columns 2—4. Added together they imply returns of 0.88 to 0.94, and we cannot reject constant returns to scale. The coefficient on labor shrinks in the final column after controlling for utilization. That change suggests labor hours are a partial proxy for utilization—not surprising, as the vast majority of labor is simply the hours of the manager (who is usually the sole worker). When the business is closed the manager likely scales down her own hours accordingly. After controlling for utilization the total returns-to-scale falls to 0.84, though we still cannot reject constant returns to scale.

Table 3: Production Elasticities (Logistical Ability Estimation)

	OLS	IV (Capital=Investment)			
	(1)	(2)	(3)	(4)	(5)
Capital	0.099**	0.470***	0.529***	0.523***	0.520***
	(0.045)	(0.157)	(0.173)	(0.181)	(0.182)
Labor	0.416***	0.413***	0.412***	0.414***	0.319***
	(0.015)	(0.015)	(0.015)	(0.016)	(0.029)
Utilization					0.025***
					(0.006)
Shops	867	862	862	853	853
Shop-Rounds	5460	5191	5191	5055	5055
rk Statistic		40	77	70	70
Sum of Elasticities	0.51	0.88	0.94	0.94	0.84
.	(0.05)	(0.16)	(0.17)	(0.18)	(0.18)
R-Squared	0.89	0.89	0.89	0.90	0.90
Fixed-Effects	Round	Round	Round	Round-Industry	Round-Industry

Note: The table reports estimates of Equation 42. Column 2 instruments for (log) capital using a dummy for whether the shop made a positive investment, and (if positive) the log of the value of the investment. Columns 3—5 instead instrument for capital using the inverse-hyperbolic sine of positive investments, minus the inverse hyperbolic sine of negative investments. All specifications also control for the restocking frequency during the round using a saturated set of dummy variables.

We extract the fixed effects from the specification in Column 4, which Equation 42 implies is an unbiased estimate of logistical ability.

5.3 Inventory Choice Ability

Equation 10 shows how we construct an estimator for inventory choice ability. As in the previous section, we need to multiply both sides by Q_{nt} to convert unobserved per-restocking quantities into observed weekly quantities. We then rescale everything by the total inventory avail-

[12] These are: Animal products, Clothing, Dried foods, Drinks, Fruits, Household products, Medicines, Personal care, Processed food, vegetables, and Miscellaneous.

able for sale during the week:

$$\frac{Q_{nt}X_{nt} - Q_{nt}\frac{Y_{nt}}{\mu_{nt}}}{Q_{nt}X_{nt}} = \frac{D}{\mathcal{I}}\frac{\Phi(\upsilon)\,\upsilon + \phi(\upsilon)}{Q_{nt}X_{nt}} \tag{43}$$

Let W_{nt} represent the left-hand side of this equation, which is the total value of unsold inventory as a share of total inventory available for sale during the week.

Assume $\mathcal{I}$ varies across shops but is fixed over time. Proposition 3 implies

$$\upsilon = \Phi^{-1}\left(\frac{mB - f}{(m+\delta)B - (1-\delta)f}\right) = \Phi^{-1}\left(\frac{m - \frac{f}{B}}{(m+\delta) - (1-\delta)\frac{f}{B}}\right)$$

which is simply a nonlinear function of the markup μ and the fixed cost of restocking as a fraction of the value of purchases $\frac{f}{B}$. Assuming δ is the same across shops, we can estimate

$$\log W_{nt} = -\log \mathcal{I}_n + h(\log X_{nt}, \log D_{nt}, \log \mu_{nt}, \log \frac{f_{nt}}{B_{nt}}, Q_{nt}) \tag{44}$$

where $h(\cdot)$ is a nonparametric function. Equation 44 can be estimated using a non-parametric fixed-effects regression, and inventory choice ability would be the following function of the estimates:

$$\log \hat{\mathcal{I}}_n = -\frac{1}{T}\sum_{t=1}^{T}\left(\log W_{nt} - \hat{h}(\log X_{nt}, \log D_{nt}, \log \mu_{nt}, \log \frac{f_{nt}}{B_{nt}}, Q_{nt}\right) \tag{45}$$

Since the coefficients from the nonparametric regression do not necessarily represent structural parameters, it is not obvious how to interpret them. Instead we estimate 5 different versions of Equation 44 to confirm that the resulting versions of inventory choice ability are similar. The versions are as follow:

1. Baseline specification—control function is linear in the variables listed in Equation 44
2. Controls for fully-interacted quadratic in the variables listed in Equation 44
3. Comparable to baseline specification except that instead of controlling for the (univariate) frequency of restocking, it controls for a vector of dummies for whether the shop restocked on any given day of the week
4. Comparable to baseline specification except that in addition to controlling for the (univariate) frequency of restocking, it controls for the days since restocking at the end of the week when unsold inventory is measured
5. Comparable to Version 4 except that instead of the (univariate and continuous) number of days since restocking, it controls for a vector of dummies for 1 day since restocking, 2 days, etc

Versions 3—5 are intended to control for any potential bias arising from how we measure unsold inventory. Taking the literal difference between purchases and sales is prone to serious measurement error. Instead we measure unsold inventory as the stock of inventory at closing time on Sunday (or the last day of the week when the shop is open, if it is closed on Sunday). By construction the stock of inventory at the end of the week is the difference between inventory purchased and inventory sold (minus inventory lost to spoilage, theft, etc.). But this measure will potentially be biased if some shops restock closer to Sunday than others even if their actual weekly restocking frequency is the same. Controlling for the date of restocking or days since restocking addresses this concern.

Table 4 presents the correlation coefficients across these five versions. The results show that all five versions are well correlated. Version 2, the quadratic estimation, is the least well-correlated with the others, but even this correlation is high (ranging from 0.85 to 0.87). In the other cases the pairwise correlation is above 0.98. Given this high correlation, in the sections that follow we use Version 1 as our estimate for inventory choice ability.

Table 4: Correlation Between Versions of Inventory Choice Ability

	Version1	Version2	Version3	Version4	Version5
Version1	1.00				
Version2	0.93	1.00			
Version3	0.98	0.91	1.00		
Version4	1.00	0.93	0.98	1.00	
Version5	1.00	0.92	0.98	1.00	1.00

Note: This table shows the correlation matrix between the five versions of inventory choice ability described in the text.

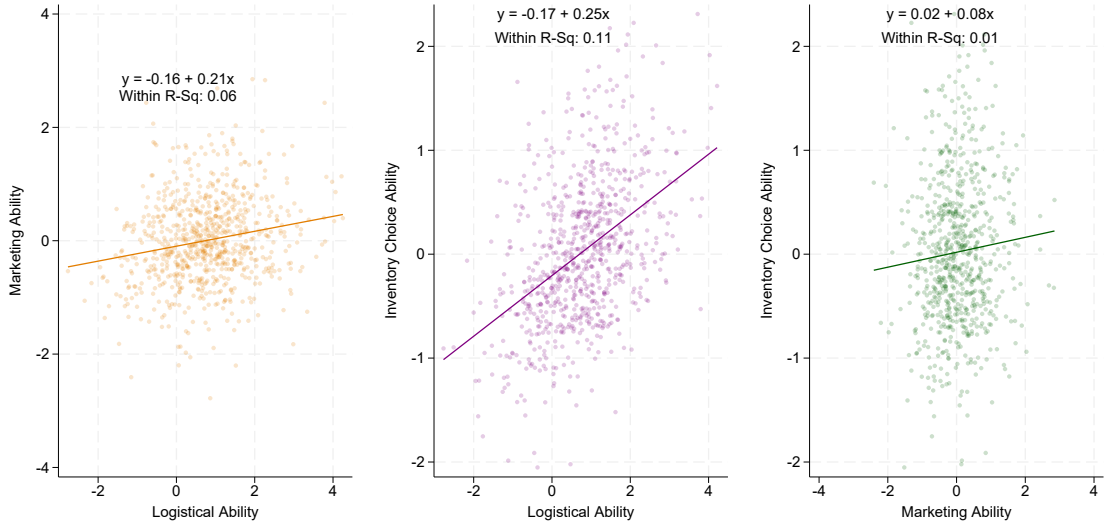
6 Results

6.1 The Three Dimensions of Ability are Independent

Section 4 shows that our model makes starkly different predictions from the standard model if the three dimensions of ability are not strongly correlated. Our first step is to establish whether this condition holds in our sample. Figure 10 shows that among our sample of retailers in Zambia the correlation between logistical and inventory choice ability is positive, but far from perfect (the correlation coefficient is 0.23). Though some of the unexplained variation may be sampling error, it is likely that there are many entrepreneurs who are highly skilled at logistics but unskilled at selecting inventory.

The simple test of whether the differences in the estimates arise from meaningful differences in ability is whether, as derived in Section 4.1, all three dimensions independently increase the level of adjusted profit. Unlike in the model, adjusted profit in the data (as defined in

Figure 10: The Correlation Between the Three Dimensions of Ability is Low



Note: Each plot presents the scatter plot, line of best fit, regression equation, and R-squared of one dimension against a different dimension. All pairwise regressions control for broad product category dummies.

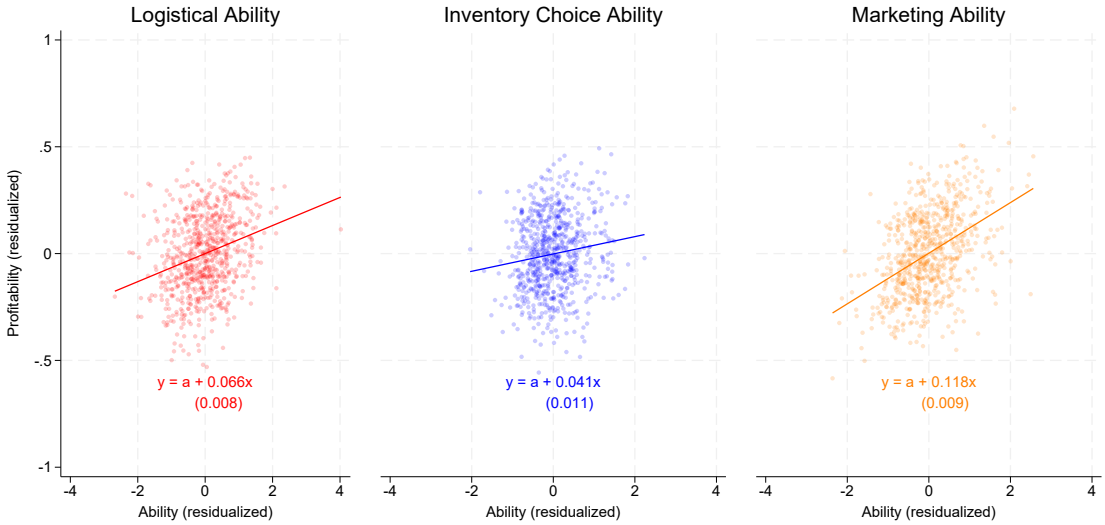
Section 2.2) can be negative, making it difficult to define a convenient rescaling transformation like the natural logarithm.[13] We instead use the cross-round percentile of adjusted profit, which gives the relative profit ranking within a round. We then calculate the within-shop average of this transformed measure across all rounds. The results are qualitatively similar if we instead use the Winsorized level of adjusted profit. Figure 11 shows partial scatter plots that plot adjusted profit against one dimension after partialing out the effect of the other dimensions and a set of broad product-type fixed-effects.[14] All measures are strongly and independently correlated with adjusted profit.

Taken together these two results imply the model's predictions will diverge from those of the standard one-stage model. One of the starkest of these is Proposition 4, which implies that a firm with higher logistical ability may actually choose a lower level of capital. Figure 12 shows that our estimates are consistent with this prediction. Logistical ability is negatively correlated with capital. The correlation with labor is also negative, though much weaker and not statistically significant. This result may reflect that labor is less of a choice than investment, as the vast majority of stores are run solely by the owner, and the most of the owner's hours are spent simply keeping the store open rather than buying inventory or stocking shelves.

[13]The inverse hyperbolic sine transformation creates an artificial bimodality in the distribution around zero, making it unsuitable.

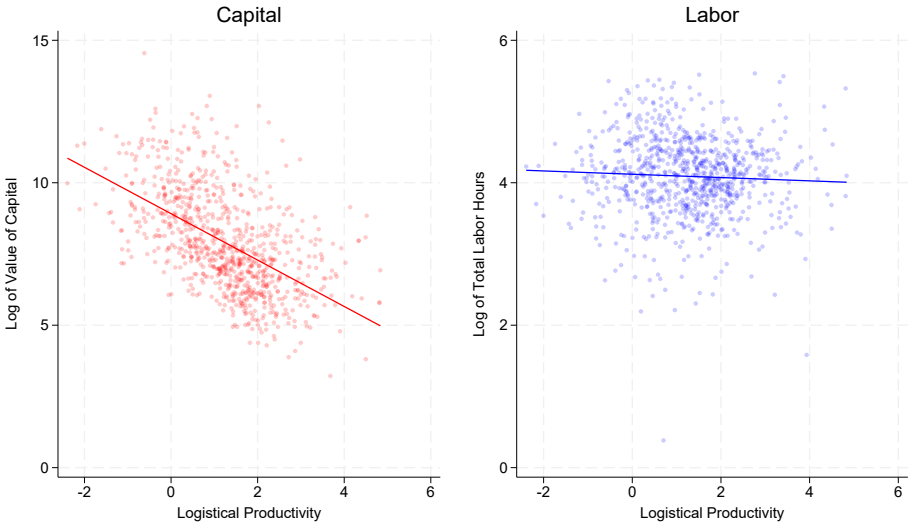
[14]The fixed effects are defined by sorting shops by their top product into the following categories: Animal, Clothing, Dried, Drinks, Fruits, Non-Food Household Products, Medicine, Personal Care, Processed Food, Vegetables,

Figure 11: All Three Dimensions of Ability are Independently Correlated with Adjusted profit



Note: Each panel shows the partial scatter plot (and line of best fit) between adjusted profit, as defined in Section 2.2, and one dimension of ability. We transform adjusted profit into its cross-round percentile (meaning a value of 1 is the highest level of profit across all shops and rounds, and 0 the lowest). Each panel partials out the other dimensions of ability and a set of product-type fixed effects. The equations below the scatter plot show the slope and standard error.

Figure 12: Capital is Negatively Correlated with Logistical Ability, as the Model Predicts



Note: We estimate logistical ability using the specification in Column 4 of Table 3, taking the store fixed-effects (estimated by the reghdfe package) as our measure of ability. The figure shows a scatter plot of our estimate against the log of capital and labor (averaged across all rounds).

6.2 The Three Measures of Ability are Better Predictors of Profit than Alternative Measures

Do the structural estimates of ability contain information about managerial ability beyond what is contained in existing measures? We answer this question with an out-of-sample test. We re-estimate the measures of ability using only the first seven rounds of data, and measure the correlation with the cross-round percentile of adjusted profit in the eighth round. Table 5 compares the predictive power of the three measures of ability to several binary measures of management suggested by McKenzie and Woodruff (2017) and standard estimate of TFP from a fixed-effects estimator of gross output (valued at cost) that instruments for capital with investment (much as we do to estimate logistical ability). In a standard Lucas Jr (1978) model, this estimate of TFP would be proportional to univariate managerial ability.

Column 1 shows that the measures of management are generally not good predictors of profitability across the shops in our sample. Of the nine measured in our survey, only three are statistically significant. One of those three (whether the shop asks customers for products they would like the shop to stock) is actually correlated with lower profitability. The nine measures of management practices together explain 3 percent of the within-product-category variation in adjusted profit. That is far below the 19 percent explained by the three measures of ability (Column 2). When combined in Column 3, the measures of management and the measures of ability together explain 22 percent of the within-product-category variation, roughly the sum of the within R-squared of the separate specifications. This last fact holds because the management practices are mostly uncorrelated with the measures of ability. The partial R-squared for the two sets of variables confirms that the three dimensions of ability account for the vast majority of the explained variation.

Column 4 shows that the standard measure of TFP explains far less of the variation than the three dimensions of ability (5 percent as compared to 19 percent). Clearly, the three dimensions together contain information beyond TFP. But Column 5 shows that logistical ability, which is the most similar to the standard TFP estimate, explains by itself slightly more variation (8 percent versus 5 percent). Given that both estimators have the same right-hand side, the left-hand side variable derived from the model drives the difference in estimates (and their predictive power). While the standard estimator uses only sales on the left-hand side, Equation 16 uses sales plus the value of unsold inventory lost to spoilage and other depreciation. This model-derived outcome evidently leads to an estimate of ability that is more informative about future profits. Finally, Column 6 runs a "horse race" regression with both measures. Of the two, logistical ability retains a positive correlation while the coefficient on TFP flips signs. This is not surprising given that they are highly correlated with one another, but logistical ability is slightly better correlated with the outcome.

---

and Miscellaneous.

Table 5: Non-Estimation Sample Correlation

	Value-Added R8					
	(1)	(2)	(3)	(4)	(5)	(6)
Visit Competitor to Compare Products, Prices	-0.002		0.005			
	(0.023)		(0.022)			
Ask Customers for Products to Stock	-0.042**		-0.034			
	(0.021)		(0.021)			
Special Offer to Attract Customers	0.016		0.008			
	(0.018)		(0.017)			
Advertise in Any Form	0.025		0.027			
	(0.023)		(0.022)			
Negotiate Lower Prices	0.001		-0.030			
	(0.021)		(0.020)			
Compare Prices of Suppliers	0.002		0.016			
	(0.024)		(0.022)			
Written Transactions	0.042**		0.029			
	(0.020)		(0.019)			
Know Cost Main Product	-0.010		0.014			
	(0.022)		(0.021)			
Written Budget	0.077***		0.059**			
	(0.029)		(0.028)			
Logistical Ability		0.048***	0.053***		0.078***	0.701***
		(0.011)	(0.011)		(0.010)	(0.059)
Inventory Choice Ability		0.050***	0.045***			
		(0.015)	(0.015)			
Marketing Ability		0.111***	0.106***			
		(0.013)	(0.013)			
Standard Measure of TFP				0.054***		-0.564***
				(0.009)		(0.053)
Shops	741	630	630	729	722	722
R-Sq Total	0.132	0.285	0.308	0.144	0.171	0.285
R-Sq Within	0.033	0.191	0.217	0.046	0.076	0.203
Partial R2, MGMT			0.032			
Partial R2, PROD			0.325			

Note: We regress the cross-round percentile of adjusted profit in Round 8 on the three measures of ability, several binary measures of management suggested by McKenzie and Woodruff (2017), and a standard fixed-effects estimate of TFP based on typical models of production. All specifications control for product-category fixed effects. We report the total R-squared, the R-squared within product-categories, and the partial R-squared for the measures of management and the three measures of ability (that is, the share of residual variation in a specification excluding that variable that would be explained if the variable were included). TFP is estimated from a simple fixed-effects estimator of gross output (valued at cost) that instruments for capital with investment (analogous to our measure of logistical ability).

6.3 A One-Stage Model Would Overestimate the Hypothetical Gains from a Reallocation of Capital

Figure 10 implies that the three dimensions of ability are only weakly correlated, making a standard model of univariate TFP a poor representation of managerial ability. Applying the standard model could yield deeply misleading conclusions for counterfactual predictions. Suppose, for example, that a researcher wanted to use a Lucas Jr (1976) style model to calculate the cost of misallocation as in Hsieh and Klenow (2009), Restuccia and Santaeulalia-Llopis (2014), or

Shenoy (2017).[15] As a basic step, the researcher might calculate the marginal product of capital

$$\beta \frac{Y_n}{K_n} \tag{46}$$

and consider the impact of reallocating 1 Zambian kwacha of capital from every shop in the bottom 50 percent to a shop in the top 50 percent of the distribution of the marginal product of capital. This is a scaled-down version of the most typical exercise in the literature of reallocating all resources optimally.

Let $\mathcal{T}$ denote the set of shops in the top 50 percent of the distribution of (46). The gains under the standard model are

$$\left(\sum_{n \in \mathcal{T}} \beta \frac{Y_n}{K_n} \right) - \left(\sum_{n \in \mathcal{T}^c} \beta \frac{Y_n}{K_n} \right) \tag{47}$$

In our model the true marginal product is $\frac{\partial \mathcal{P}_n}{\partial K_n}$. To simplify the calculations we apply the expressions derived in Section 4.1 for the model where the share of unsold inventory lost equals 1 ($\delta = 1$). In this simplified model, the marginal product is given by:

$$\frac{\partial \mathcal{P}_n}{\partial K_n} = \beta \frac{f_n}{B_n} \frac{Q_n X_n}{K_n} \tag{48}$$

which alters the expression from the standard model by replacing sales with the total quantity of inventory that the shop wants to make available for sale during the period, and rescales the expression by the fixed cost per restocking (normalized by the quantity bought per restocking). This expression clarifies that additional capital lets the firm buy more per restocking, allowing it to display the same quantity of goods while paying the fixed cost less frequently.

Let $\bar{\mathcal{T}}$ denote the set of shops in the top 50 percent of the distribution of (48). Then the hypothetical gains in our model are

$$\left(\sum_{n \in \bar{\mathcal{T}}} \beta \frac{f_n}{B_n} \frac{Q_n X_n}{K_n} \right) - \left(\sum_{n \in \bar{\mathcal{T}}^c} \beta \frac{f_n}{B_n} \frac{Q_n X_n}{K_n} \right) \tag{49}$$

We calculate Equations 47 and 49 in our dataset of Zambian retailers. The hypothetical gains are 3.4 percent of total sales under the standard model, more than eight times as large as the 0.4 percent hypothetical gains in our model.[16] The gap in gains from reallocation is largely driven by the difference across each model's assumptions about the impact of capital. The standard model assumes capital directly raises sales, whereas in our model it increases profit by reducing

[15] Though Hsieh and Klenow (2009) uses a Melitz-style model instead of a Lucas-style model, for the purposes of calculating misallocation the two approaches are isomorphic.

[16] We calculate the gains as a share of total sales to keep the denominator consistent, though in our model adjusted profit is the better measure of output.

the cost of sourcing inventory. Since efficiently sourcing inventory is only one part of the firm's operation, the misallocation of capital is simply less important.

6.4 A Hypothetical Intervention Based on the Standard Model Would be Mistargeted

Consider a policy that gives a small capital grant to firms with a high marginal product of capital. To be precise, suppose the government ranks firms by the standard measure and gives the top 50 percent a small quantity ΔK of physical capital. For simplicity we assume the grant is small enough that we can approximate the impact on aggregate profit with

$$\sum_{n \in \mathcal{T}} \frac{\partial \mathcal{P}_n}{\partial K_n} \Delta K \tag{50}$$

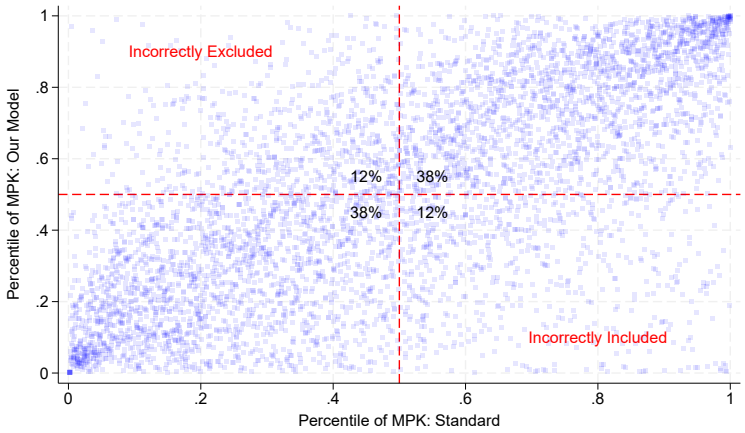
A government using the standard model would target the set of firms in $\mathcal{T}$, as defined in the previous section using Equation 46. But it should have targeted the firms in $\bar{\mathcal{T}}$ using Equation 48. Figure 13 shows where each shop-round falls in the distribution of the marginal product of capital based on both the standard model and our model. The top left quadrant is $\mathcal{T}^c \cap \bar{\mathcal{T}}$, shops excluded from the policy that should have been included. The bottom right quadrant is the opposite, $\mathcal{T} \cap \bar{\mathcal{T}}^c$, shops that were included and should have been excluded. Roughly 24 percent of the sample is either included or excluded erroneously when the grant is targeted using the standard model. We calculate that the total benefits of the grant would be 11.5 percent higher if it were targeted correctly.

An even worse form of mistargeting, however, lies in the specific intervention chosen for each firm. In our multi-stage model it is not obvious whether capital is the most valuable resource for every shop. Consider instead a program that gives 1 ZMW of fixed advertising investment $\bar{a}$. Section 5.1 estimated marketing ability by defining $\bar{a}$ to be the inverse hyperbolic sine of the value of signs during the baseline survey. Let $\bar{A}$ be the level of fixed advertising investment at baseline. Then Equation 27 implies that the change in profit from 1 additional ZMW of advertising investment is

$$\begin{aligned}
\Delta \mathcal{P} &= \left(e^{\omega \delta_1 \operatorname{arcsin}^{-1}(\bar{A}+1)} - e^{\omega \delta_1 \operatorname{arcsin}^{-1}(\bar{A})} \right) \left\{ \mathcal{M}(Z)\zeta - \frac{\mu}{\mathcal{I}} \phi \left[\Phi^{-1} \left(\frac{\mathcal{M}(Z)}{\mu} \right) \right] \right\} \\
&= \left(e^{\omega \delta_1 \operatorname{arcsin}^{-1}(\bar{A}+1)} - e^{\omega \delta_1 \operatorname{arcsin}^{-1}(\bar{A})} \right) e^{-\omega \delta_1 \operatorname{arcsin}^{-1}(\bar{A})} \mathcal{P}
\end{aligned} \tag{51}$$

For each shop-round in our sample we calculate the relative impact of 1 ZMW of advertising revenue versus 1 ZMW of physical capital. To be precise, we subtract the marginal product of capital (Equation 48) from Equation 51. Figure 14 plots the histogram of the relative impact (after Winsorizing the top and bottom 1 percent). The figure shows that the difference is positive

Figure 13: Marginal Product of Capital: Standard Versus Current Model



Note: We assign each shop-round its percentile in the distribution of the marginal product of capital under two models: the standard model (46) and our model (48). The scatter plot shows the relationship between these percentiles. The dashed lines show the cutoff for being targeted by a capital grant given to shops in the top 50 percent by each metric. The numbers show the share of the total sample that fall into each quadrant.

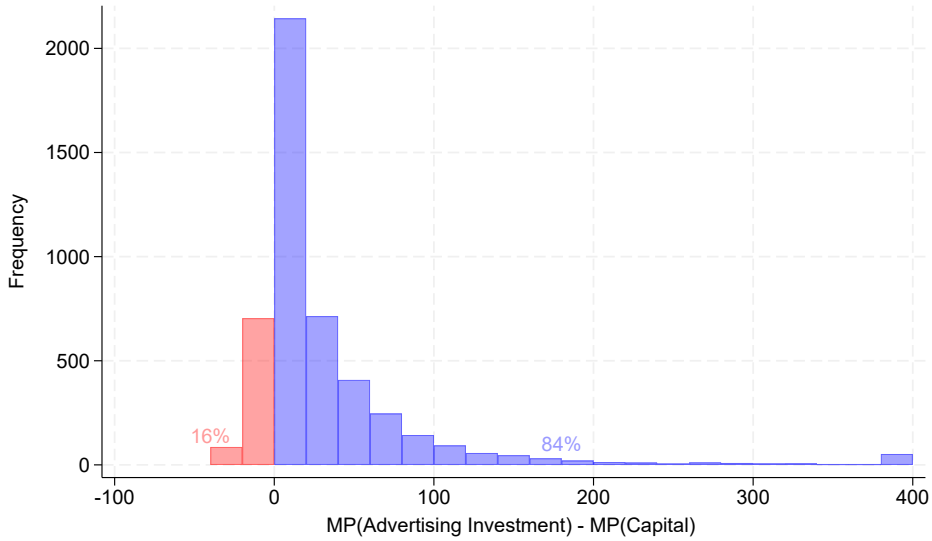
for 84 percent of shop-rounds, implying the additional advertising is more valuable for the vast majority of the sample.

One caveat to this calculation is that it assumes the targeted shop(s) are a small share of the total sector, which allows us to ignore that one shop's advertising investment may partly cancel out that of another. Formally, we can assume the denominator of Equation 20 does not change. This is the appropriate counterfactual for either a small intervention or one targeted sparsely over a large area. Figure 14 implies that an intervention supporting retailer advertising would lead to higher profit gains—for some shops, many times greater—than one that distributed equipment.

7 Conclusion and Directions for Future Research

This paper proposes a three-stage model of retail production where success in each stage requires a distinct form of entrepreneurial ability. We use the model to estimate the three dimensions of ability in a novel dataset collected from small retailers in Lusaka, Zambia. Our estimates are more strongly correlated with future value-added than survey-based measures of management practices and a traditional fixed-effects estimate of TFP. The three dimensions of ability are only weakly correlated, which we show implies that a standard one-stage model could yield misleading conclusions in counterfactual analysis and suboptimal results if used to guide an intervention. We show in our dataset that the standard model would dramatically

Figure 14: Relative Impact of 1 Unit of Fixed Advertising Investment (Signs) Versus 1 Unit of Capital



Note: We subtract the marginal product of capital from the marginal product of fixed advertising investment. The histogram plots this difference (each observation is a shop-round). The percentages show the share of observations where the difference is positive (shaded blue) and negative (shaded red). A positive value implies that 1 dollar of advertising would raise profit by more than 1 dollar of capital.

overestimate the impact of a hypothetical capital reallocation. Targeting capital grants using the standard model would mis-target nearly one-quarter of the sample and fail to recognize that the vast majority of the sample would benefit more from a grant of advertising investment.

Our results imply that the standard one-stage model of production can yield misleading conclusions. Though we argue the limitations of that model are especially clear in retail, they can also apply to other sectors. While a manufacturer does not choose inventory, for example, it does have to choose inputs whose productive value may be imperfectly observed. If some entrepreneurs are skilled at managing a production line, but less skilled at sourcing high-quality inputs, there is a potential for the one-stage model to give misleading results for manufacturing, too. Designing industry-specific, multi-stage models of production may help solve at least part of the mystery of why some firms are far more successful than others and provide valuable insights into the roles that entrepreneurs play in driving those differences.

References

ACKERBERG, D., C. L. BENKARD, S. BERRY, AND A. PAKES (2007): "Econometric Tools for Analyzing Market Outcomes," *Handbook of Econometrics*, 6, 4171–4276.

ADHVARYU, A., J.-F. GAUTHIER, A. NYSHADHAM, AND J. TAMAYO (2024): "Absenteeism, productivity, and relational contracts inside the firm," *Journal of the European Economic Association*, 22, 1628–1677.

ADHVARYU, A., A. NYSHADHAM, AND J. TAMAYO (2022): "Managerial quality and productivity dynamics," *The Review of Economic Studies*, 90, 1569–1607.

ANDERSON, S. J., R. CHANDY, AND B. ZIA (2018): "Pathways to profits: The impact of marketing vs. Finance skills on business performance," *Management science*, 64, 5559–5583.

BASSI, V., J. H. LEE, A. PETER, T. PORZIO, R. SEN, AND E. TUGUME (2023): "Self-employment within the firm," *SSRN Electronic Journal*.

BERNARD, A. B., E. DHYNE, G. MAGERMAN, K. MANOVA, AND A. MOXNES (2022): "The origins of firm heterogeneity: A production network approach," *The journal of political economy*, 130, 1765–1804.

BLOOM, N., A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2010): "Why do firms in developing countries have low productivity?" *American Economic Review*, 100, 619–623.

BOEHM, J. AND E. OBERFIELD (2020): "Misallocation in the Market for Inputs: Enforcement and the Organization of Production," *The quarterly journal of economics*.

BROOKS, W., K. DONOVAN, AND T. R. JOHNSON (2018): "Mentors or Teachers? Microenterprise Training in Kenya," *American Economic Journal. Applied Economics*, 10, 196–221.

COLLARD-WEXLER, A. AND J. DE LOECKER (2016): "Production Function Estimation With Measurement Error in Inputs," Tech. rep., National Bureau of Economic Research.

EATON, J. AND S. KORTUM (2002): "Technology, Geography, and Trade," *Econometrica: Journal of the Econometric Society*, 70, 1741–1779.

EHRLICH, I. AND L. FISHER (1982): "The Derived Demand for Advertising: A Theoretical and Empirical Investigation," *The American Economic Review*, 72, 366–388.

HSIEH, C. AND P. KLENOW (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, 1403–1448.

INKLAAR, R., R. MARAPIN, AND K. GRÄLER (2023): "Tradability and sectoral productivity differences across countries," .

LAGAKOS, D. (2016): "Explaining Cross-Country Productivity Differences in Retail Trade," *The Journal of Political Economy*, 124, 579–620.

LUCAS JR, R. (1978): "On the Size Distribution of Business Firms," *The Bell Journal of Economics*, 9, 508–523.

LUCAS JR, R. E. (1976): "Econometric Policy Evaluation: A Critique," in *Carnegie-Rochester conference series on public policy*, Elsevier, vol. 1, 19–46.

MACCHIAVELLO, R. AND A. MORJARIA (2021): "Competition and relational contracts in the Rwanda coffee chain," *The Quarterly Journal of Economics*, 136, 1089–1143.

MCKENZIE, D. AND C. WOODRUFF (2017): "Business Practices in Small Firms in Developing Countries," *Management Science*, 63, 2967–2981.

QUINN, S. AND C. WOODRUFF (2019): "Experiments and Entrepreneurship in Developing Countries," *Annual Review of Economics*.

RESTUCCIA, D. AND R. ROGERSON (2008): "Policy Distortions and Aggregate Productivity with Heterogeneous Establishments," *Review of Economic Dynamics*, 11, 707–720.

RESTUCCIA, D. AND R. SANTAEULALIA-LLOPIS (2014): "Land Misallocation and Productivity," .

SHENOY, A. (2017): "Market Failures and Misallocation," *Journal of Development Economics*, 130, 65–80.

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

A Proofs of Propositions in the Main Text

A.1 Derivation of (Expected) Total Sales

Substitute Equation 8 into the expression for expected sales:

$$\mathbb{E}[Y] = \sum_{j=1}^{J} P_j \left\{ \left[1 - \Phi\left(\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J} \right) \right] x_j + \Phi\left(\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J} \right) \mathbb{E}[\alpha_j DS/J \mid \alpha_j DS/J < x_j] \right\} \tag{52}$$

$$= \mu \left(\sum_{j=1}^{J} \kappa_j x_j \right) [1 - \Phi(\upsilon)] + \mu \Phi(\upsilon) \left(\sum_{j=1}^{J} \kappa_j \mathbb{E}[\alpha_j DS/J \mid \alpha_j DS/J < x_j] \right) \tag{53}$$

$$= \mu \left(\sum_{j=1}^{J} \kappa_j x_j \right) [1 - \Phi(\upsilon)] + \mu \left(\frac{1}{J} \sum_{j=1}^{J} \kappa_j \tilde{\alpha}_j DS \right) \Phi(\upsilon) - \mu\sigma DS \left(\frac{1}{J} \sum_{j=1}^{J} \kappa_j \right) \phi(\upsilon) \tag{54}$$

$$= \mu X [1 - \Phi(\upsilon)] + \mu\zeta DS\Phi(\upsilon) - \mu\sigma DS\dot{\kappa}\phi(\upsilon) \tag{55}$$

where (53) follows from (8) and the fixed markup assumption, (54) follows from the truncated expectation of a normal random variable, and (55) from applying (3) and (6). Collect terms and apply the definition of υ to form a final expression for sales within a restocking period:

$$\mathbb{E}[Y] = \mu \Big\{ X - \sigma\dot{\kappa} DS [\Phi(\upsilon)\upsilon + \phi(\upsilon)] \Big\} \tag{56}$$

To show the final result, observe that for any j,

$$Var(P_j \min\{x_j, DS\alpha_j/J\}) \leq Var(P_j DS\alpha_j/J) = \frac{P_j^2 D^2 S^2 \sigma^2}{J^2}$$

because x_j is a known constant. Then

$$\lim_{J\to\infty} \sum_{j=1}^{J} Var(P_j \min\{x_j, DS\alpha_j/J\}) < \infty$$

for any fixed X, D, S. Then the Kolmogorov two-series theorem implies that

$$0 = \sum_{j=1}^{J} [P_j \min\{x_j, DS\alpha_j/J\} - \mathbb{E}(P_j \min\{x_j, DS\alpha_j/J\})] \tag{57}$$

$$= \sum_{j=1}^{J} [P_j \min\{x_j, DS\alpha_j/J\}] - \mathbb{E}\left(\sum_{j=1}^{J} P_j \min\{x_j, DS\alpha_j/J\} \right) \tag{58}$$

$$= Y - \mathbb{E}[Y] \tag{59}$$

will converge in $\Re$ as $J \to \infty$.

A.2 Proof that One-Period Optimization Coincides with Multi-period Optimization

We can generalize the static problem to a dynamic problem where the shop chooses $\{x_j\}$ at period t anticipating that it can restock after an interval of length S_t, which we assume is either pre-determined or will be chosen based on information known currently (and is thus not a random variable).

Define

$$U_{t+S_t} = \sum_{j=1}^{J} \kappa_{jt} \max\{x_{jt} - D_t S_t \alpha_{jt}/J, 0\}$$

which is the replacement cost of inventory unsold at the end of the period. Under the assumption that the store will trade in unsold inventory and retain a share $\mathcal{D}(S_t)$ that does not depend on j, the value of the trade-in is $\mathcal{D}(S_t)U_{t+S_t}$. This term will, together with the firm's logistical capacity B_{t+S_t}, determine the upper bound of the expenditure constraint on new inventory.

The value function at time t is

$$\mathcal{V}(U_t; \{\alpha_{jt}\}, \{P_{jt}\}) \tag{60}$$

$$= \max_{\{x_{jt}\}} \mathbb{E}\Big[\sum_{j=1}^{J} P_{jt} \min\{x_{jt}, D_t S_t \alpha_{jt}/J\} + \vartheta_t \left(\mathcal{D}(S_{t-1})U_t + B_t - \sum_{j=1}^{J} \kappa_{jt} x_{jt} \right) \tag{61}$$

$$+ \mathcal{V}(U_{t+S_t}; \{\alpha_{j,t+S_t}\}, \{P_{j,t+S_t}\}) \Big] \tag{62}$$

where we suppress several predetermined inputs to the value function (e.g. D_t, S_t, S_{t-1}) to keep the notation concise. As in the main text, ϑ_t is a Kuhn-Tucker multiplier

The first-order condition is

$$\kappa_{jt}\vartheta_t = P_{jt}\left[1 - \Phi\left(\frac{x_{jt} - D_t S_t \tilde{\alpha}_{jt}/J}{\sigma D_t S_t/J}\right)\right] + \mathbb{E}\left[\mathcal{V}_1(U_{t+S_t}; \{\alpha_{j,t+S_t}\}, \{P_{j,t+S_t}\}) \frac{\partial U_{t+S_t}}{\partial x_{jt}}\right] \tag{63}$$

and the envelope condition with respect to X_t implies

$$\mathcal{V}_1(U_{t+S_t}; \{\alpha_{j,t+S_t}\}, \{P_{j,t+S_t}\}) = \vartheta_{t+S_t}\mathcal{D}(S_t)$$

while

$$\frac{\partial U_{t+S_t}}{\partial x_{jt}} = \kappa_{jt}\mathbb{I}(x_{jt} \geq D_t S_t \alpha_{jt}/J)$$

Combine these conditions and apply the fixed markup assumption:

$$\vartheta_t = \mu_t \left[1 - \Phi\left(\frac{x_{jt} - D_t S_t \tilde{\alpha}_{jt}/J}{\sigma D_t S_t/J}\right)\right] + \mathcal{D}(S_t)\Phi\left(\frac{x_{jt} - \tilde{\alpha}_{jt} D_t S_t/J}{\sigma D_t S_t/J}\right)\mathbb{E}[\vartheta_{t+S_t} \mid x_{jt} \geq D_t S_t \alpha_{jt}/J] \tag{64}$$

This condition will be satisfied for any choice of x_{jt} that makes the right-hand side of Equation 64 equal for all j.

Suppose that

$$x_{jt} = \tilde{\alpha}_{jt} D_t S_t/J + (\sigma D_t S_t/J)\upsilon_t$$

for some quantity υ_t. Then

$$\vartheta_t = \mu_t \left[1 - \Phi\left(\upsilon_t\right)\right] + \mathcal{D}(S_t)\Phi\left(\upsilon_t\right)\mathbb{E}\left[\vartheta_{t+S_t} \mid \upsilon_t \geq \frac{\alpha_{jt} - \tilde{\alpha}_{jt}}{\sigma}\right] \tag{65}$$

If J is very large, then

$$\mathbb{E}\left[\vartheta_{t+S_t} \mid \upsilon_t \geq \frac{\alpha_{jt} - \tilde{\alpha}_{jt}}{\sigma}\right] \approx \mathbb{E}\left[\vartheta_{t+S_t}\right]$$

because the unsold inventory of any one good j will be too small a share of U_{t+S_t} to materially impact its value, and the independence of the noise terms $u_{jt} = \frac{\alpha_{jt} - \tilde{\alpha}_{jt}}{\sigma}$ ensures they are uninformative about other goods. In the limit as $J \to \infty$ the aggregate unsold inventory U_{t+S_t} converges to a constant, and ϑ_{t+S_t} is a constant. Then the first-order condition is satisfied.

Now define

$$X_t = \sum_{j=1}^{J} \kappa_{jt} x_{jt}$$

Then simple arithmetic suffices to show that

$$\upsilon_t = \frac{X_t - \zeta_t D_t S_t}{\sigma D S \dot{\kappa}_t}$$

where ζ_t and $\dot{\kappa}_t$ are defined as in the main text. Then choosing the individual $\{x_{jt}\}$ is equivalent to choosing X_t and setting $x_{jt} = \tilde{\alpha}_{jt} DS/J + (\sigma DS/J)\upsilon_t$. We have proven that the choices of x_{jt} in this dynamic problem coincide with those of the simple static problem.

As an aside, simple algebra shows that the constraint in Equation 62 is equivalent to

$$\mathcal{D}(S_{t-1})U_t + B_t - \sum_{j=1}^{J} \kappa_{jt} x_{jt} = \mathcal{D}(S_{t-1})(X_t - Y_t/\mu_t) + B_t - X_t \tag{66}$$

$$\Rightarrow X_t \leq Z_t K_t^{\beta} L_t^{\nu} + \mathcal{D}(S_{t-1})(X_t - Y_t/\mu_t) \tag{67}$$

which is simply the constraint (15) in Section 3.2 when $\mathcal{D}(S_{t-1}) = 1 - \delta$.

A.3 Proof of Proposition 2

Let ϑ be the Lagrange multiplier on the inequality constraint (15). Define

$$G = X - (1-\delta)(X - Y/\mu) \tag{68}$$

$$= X - (1-\delta)\frac{\sigma D\dot{\kappa}}{Q}\left[\Phi\left(\upsilon\right)\upsilon + \phi(\upsilon)\right] \tag{69}$$

which is simply the new spending on inventory (beyond the undepreciated existing stock).

Then the Lagrangian is

$$\mathcal{L} = Q\left\{Y - G - f\right\} + \vartheta\Big[ZK^{\beta}L^{\nu} - G\Big] \tag{70}$$

The first-order conditions are

$$\vartheta\frac{\partial G}{\partial X} = \left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]Q \tag{71}$$

$$\vartheta\frac{\partial G}{\partial Q} = \left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f \tag{72}$$

Now suppose that the constraint does not bind and $\vartheta = 0$. Then Equation 71 becomes

$$0 = \Big[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + \left\{1-(1-\delta)\Phi\left(\upsilon\right)\right\}\Big]Q$$

The equation is satisfied if and only if $Q = 0$ or the term in square brackets is zero. Since $Q > 0$, it must be that the term in square brackets is zero. But then Equation 72 implies that $f = 0$, a contradiction. Then it must be that the constraint always binds.

A.4 Proof of Proposition 3

Divide (72) by (71) and rearrange:

$$\frac{\frac{\partial G}{\partial Q}}{\frac{\partial G}{\partial X}} = \frac{\left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f}{\left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]Q} \tag{73}$$

$$\frac{(1-\delta)\left\{Y/\mu - X + \Phi\left(\upsilon\right)X\right\}}{(1-\delta)\Phi\left(\upsilon\right) - 1} = \frac{\left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f}{\left[\mu\left\{1-\Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]} \tag{74}$$

$$\frac{ZK^{\beta}L^{\nu} + \Big[(1-\delta)\Phi\left(\upsilon\right) - 1\Big]X}{(1-\delta)\Phi\left(\upsilon\right) - 1} = X - \frac{f}{\mu - 1 - \left\{\mu - (1-\delta)\right\}\Phi\left(\upsilon\right)} \tag{75}$$

$$X - \frac{ZK^{\beta}L^{\nu}}{1-(1-\delta)\Phi\left(\upsilon\right)} = X - \frac{f}{\mu - 1 - \left\{\mu - (1-\delta)\right\}\Phi\left(\upsilon\right)} \tag{76}$$

$$\frac{f}{\mu\left[1-\Phi\left(\upsilon\right)\right] - \left[1-(1-\delta)\Phi\left(\upsilon\right)\right]} = \frac{ZK^{\beta}L^{\nu}}{1-(1-\delta)\Phi\left(\upsilon\right)} \tag{77}$$

Define $m = \mu - 1$ and $B = ZK^{\beta}L^{\nu}$.

$$\frac{m-(m+\delta)\Phi(\upsilon)}{f} = \frac{1-(1-\delta)\Phi(\upsilon)}{B} \tag{78}$$

$$mB-(m+\delta)B\Phi(\upsilon) = f-(1-\delta)f\Phi(\upsilon) \tag{79}$$

$$mB-f = [(m+\delta)B-(1-\delta)f]\Phi(\upsilon) \tag{80}$$

$$\Phi^{-1}\left(\frac{mB-f}{(m+\delta)B-(1-\delta)f}\right) = \upsilon \tag{81}$$

A.5 Proof of Proposition 4

Suppose the firm gets to choose its capital and labor given rental prices p^K and p^L . The firm would solve

$$\max_{Q^*,X^*} Q^*\left\{Y(X^*,Q^*)-\left[X^*-(1-\delta)(X^*-Y(X^*,Q^*)/\mu)\right]-f\right\} \tag{82}$$

$$+\vartheta[(1-\delta)(X^*-Y(X^*,Q^*)/\mu)+ZK^{\beta}L^{\nu}-X^*]-p^KK-p^LL \tag{83}$$

where X^* and Q^* satisfy the conditions derived in the previous section. By the Envelope Theorem, the optimal choice of K satisfies

$$\vartheta\beta\frac{ZK^{\beta}L^{\nu}}{K} = p^K \tag{84}$$

Subbing (71) into (72) shows that

$$-\vartheta\frac{\partial G}{\partial Q} = [\mu\{1-\Phi(\upsilon)\}+(1-\delta)\Phi(\upsilon)-1]\,Q\frac{X}{Q}-f \tag{85}$$

$$= \left(-\vartheta\frac{\partial G}{\partial X}\right)\frac{X}{Q}-f \tag{86}$$

$$\Rightarrow \vartheta = \frac{f}{\frac{\partial G}{\partial Q}-\frac{\partial G}{\partial X}\frac{X}{Q}} \tag{87}$$

Take the partial derivative of G with respect to Q and substitute in the (binding) constraint:

$$\frac{\partial G}{\partial Q} = (1-\delta)\frac{1}{Q}\{Y/\mu-X+\Phi(\upsilon)X\} \tag{88}$$

$$= \frac{1}{Q}\{(1-\delta)(Y/\mu-X)+X-X+(1-\delta)\Phi(\upsilon)X\} \tag{89}$$

$$= \frac{1}{Q}\left\{ZK^{\beta}L^{\nu}+((1-\delta)\Phi(\upsilon)-1)X\right\} \tag{90}$$

$$= \frac{1}{Q}\left\{ZK^{\beta}L^{\nu}+\frac{\partial G}{\partial X}X\right\} \tag{91}$$

Subbing this back into (87) yields

$$\vartheta = \frac{fQ}{ZK^{\beta}L^{\nu}} \tag{92}$$

which substituted into (84) yields

$$K = \beta \frac{fQ}{p^K} \tag{93}$$

Then

$$\frac{\partial K}{\partial Z} \propto \frac{\partial Q}{\partial Z}$$

To sign this derivative, we argue as follows:

1. A small increase in Z creates slack in the intertemporal inventory constraint
2. As proven earlier, the shop will always adjust X and Q to ensure the constraint binds
3. To reduce slack in the constraint, the shop must increase X and reduce Q
4. Thus, $\frac{\partial Q}{\partial Z} \leq 0$

1 follows from the constraint. Define the slackness in the constraint as

$$\tilde{G} = (1-\delta)(X - Y/\mu) + ZK^{\beta}L^{\nu} - X$$

where $\tilde{G} \geq 0$. Clearly $\tilde{G}$ is increasing in Z.

2 has already been proven

To see 3: first note that

$$\frac{\partial \tilde{G}}{\partial X} = \frac{\partial G}{\partial X} = (1-\delta)\Phi(\upsilon) - 1 < 0$$

Hence, increasing X reduces slack in the constraint. Meanwhile,

$$\frac{\partial \tilde{G}}{\partial Q} = -\frac{\partial Y/\mu}{\partial Q}$$

By definition

$$Y = \int_0^1 P_j \min\{x_j, DS\dot{\alpha}_j\}\, dj$$

Since $S = 1/Q$ it is clear that an increase in Q reduces Y , implying $\frac{\partial \tilde{G}}{\partial Q} > 0$. Increasing Q creates slack in the constraint, while reducing it reduces slack.

With 1—3 shown, 4 must follow, and with it the proposition.

B Extensions to the Model

B.1 Product-Level Sales Depend on Prices

Suppose that customer-level demand for good j is decreasing in the price:

$$\frac{\alpha_j}{J} P_j^{-\mathcal{K}}$$

where $\mathcal{K}$ is an unknown but common parameter across firms and customers. Then Equation 5 becomes

$$x_j = \frac{\tilde{\alpha}_j DS}{J} P_j^{-\mathcal{K}} + \frac{\sigma DS}{J} P_j^{-\mathcal{K}} \Phi^{-1} \left(1 - \frac{\lambda}{\mu}\right) \tag{94}$$

Modify the definitions of ζ and $\dot{\kappa}$ to

$$\zeta = \frac{1}{J} \sum_{j=1}^{J} \kappa_j \tilde{\alpha}_j P_j^{-\mathcal{K}}$$

$$\dot{\kappa} = \frac{1}{J} \sum_{j=1}^{J} \kappa_j^{1-\mathcal{K}}$$

The interpretation of ζ, the total expected per-customer demand (valued at cost), remains unchanged. But $\dot{\kappa}$ is now a CES price aggregator, where the demand curve of customers determines the basket of goods.

Redefine

$$\upsilon = \frac{X - \zeta DS}{\sigma DS \dot{\kappa}} \mu^{-\mathcal{K}} \tag{95}$$

Then algebra parallel to that in the main text yields

$$Y \approx \mu \left\{ X - \sigma \dot{\kappa} DS \mu^{-\mathcal{K}} \left[\Phi(\upsilon) \upsilon + \phi(\upsilon) \right] \right\} \tag{96}$$

as J becomes large. Subbing Equation 96 into the logistical productivity problem yields first-order conditions identical to those in the main text. Calculations identical to those in Appendix A.4 yield

$$\upsilon = \Phi^{-1} \left(\frac{mB - f}{(m+\delta)B - (1-\delta)f} \right) \tag{97}$$

Subbing Equation 97 into Equation 96 and solving for W_{nt} as defined in Section 5 implies a nonparametric estimating equation identical to Equation 44. The only difference is that in the final measure of inventory choice productivity

$$\hat{\imath}_n \approx \log \sigma + \log \dot{\kappa}$$

the average cost $\dot{\kappa}$ is an order price aggregator rather than the simple unweighted average of order prices.

B.2 Depreciation Increases with Period Length

The basic model in the main text assumed a constant fraction δ of unsold inventory was lost at each restocking regardless of the time between restocking S. This appendix extends the model to allow total depreciation to vary with period length (or equivalently, with the frequency of restocking Q).

Let $\mathcal{D}(Q) = \mathcal{D}(1/S)$ be the *undepreciated* share of unsold inventory. We assume $\mathcal{D}'(Q) \geq 0$, meaning the undepreciated share increases as the restocking frequency increases (and the time between restocking decreases). For example,

$$\mathcal{D}(Q) = e^{-\frac{\delta}{Q}} = e^{-\delta S}$$

is standard exponential depreciation. The model in the main text is the special case $\mathcal{D}(Q) = 1 - \delta$.

We modify the optimization over the restocking plan as

$$\max_{Q,X} Q\left\{Y - \Big[X - (1-\delta)(X - Y/\mu)\Big] - f\right\} - \upsilon(H)$$

Subject to:

$$X \leq \mathcal{D}(Q)(X - Y/\mu) + HZK^{\beta}L^{\nu} \tag{98}$$

which replaces the constraint (15) with (98). Aside from using the general restocking-dependent depreciation function, this equation adds a new term to logistical capacity: utilization or "hours" H. This term reflects the hours that the shop is open, which increases the restocking capacity at a cost $\upsilon(H)$ which is increasing and now appears in the objective function.

The new first-order conditions are

$$\vartheta\frac{\partial G}{\partial X} = \left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + \mathcal{D}(Q)\Phi\left(\upsilon\right) - 1\right]Q \tag{99}$$

$$\vartheta\frac{\partial G}{\partial Q} = \left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + \mathcal{D}(Q)\Phi\left(\upsilon\right) - 1\right]X - f + \mathcal{D}'(Q)\sigma\dot{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big] \tag{100}$$

The proof of Proposition 2 from Appendix A.3 no longer holds because the term in (100) with $\mathcal{D}'(Q)$ is positive, implying that there may be an interior solution to the optimization. The reason is that the shop may now find it optimal to restock more frequently than is strictly necessary and hold a smaller inventory than it otherwise could because doing so reduces the value of inventory lost to depreciation.

But the addition of hours H to the capacity term does guarantee that the constraint will bind. The first-order condition with respect to hours is

$$\vartheta ZK^{\beta}L^{\nu} = \upsilon'(H)$$

This equation cannot be satisfied if $\vartheta = 0$. Intuitively, if the shop has spare restocking capacity, it will simply reduce its utilization (operate for fewer hours) until the constraint just binds. That is, the firm will either be constrained at full capacity, or it will achieve a global maximum at an interior solution and will reduce capacity until the constraint just binds at the global maximum.

We can remove the Lagrange multiplier from the first-order conditions by dividing (100) by (99). This ratio yields the equation

$$\frac{f - \mathcal{D}'(Q)\sigma\dot{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big]}{\mu\left[1 - \Phi\left(\upsilon\right)\right] - \left[1 - \mathcal{D}(Q)\Phi\left(\upsilon\right)\right]} = \frac{B + \mathcal{D}'(Q)\sigma\dot{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big]}{1 - \mathcal{D}(Q)\Phi\left(\upsilon\right)} \tag{101}$$

where $B = HZK^{\beta}L^{\nu}$ This expression reduces to (77) when $\mathcal{D}'(Q) = 0$ as in the main text.

Define the elasticity of $\mathcal{D}$ with respect to Q

$$\mathcal{E}(Q) = \frac{\mathcal{D}'(Q)Q}{\mathcal{D}(Q)}$$

With some rearrangement (101) reduces to

$$\frac{\mu\Big[1 - \Phi(\upsilon)\Big]}{1 - \mathcal{D}(Q)\Phi(\upsilon)} = \frac{f + B}{B + \mathcal{E}(Q)\Big[X - B\Big]} \tag{102}$$

This equation has a solution

$$\upsilon = \bar{\upsilon}(f, B, Q, X, \mu)$$

for some function $\bar{\upsilon}$. This expression can be substituted into Equation 9 to yield the same non-parametric estimating equation as derived in Section 3.

B.3 Dynamic Relationship Capital

$$d_{nt} = \omega r_{nt} - \omega\left[G\tilde{p}_{nt} + (1 - G)\hat{p}_n\right] + \phi_{gt} \tag{103}$$

$$= \omega\Big(\bar{r}_n + \delta a_{nt} + \rho r_{n,t-1} + e_{nt}\Big) - \omega\left[G\tilde{p}_{nt} + (1 - G)\hat{p}_n\right] + \phi_{gt} \tag{104}$$

The signal $\tilde{p}_{nt}$ is not observed but can be replaced by its definition in (18).

$$d_{nt} = \omega r_{nt} - \omega\left[G\bar{p}_{nt} + (1 - G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{105}$$

$$= \omega\Big(\bar{r}_n + \delta a_{nt} + \rho r_{n,t-1} + e_{nt}\Big) - \omega\left[G\bar{p}_{nt} + (1 - G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{106}$$

Combine (106) with a one-period lag of (105) :

$$d_{nt} = \omega\Big(\bar{r}_n + \delta a_{nt} + e_{nt}\Big) + \rho\Big(d_{n,t-1} + \omega\left[G\bar{p}_{n,t-1} + (1-G)\hat{p}_n + G\nu_{n,t-1}\right] - \phi_{g,t-1}\Big) \tag{107}$$

$$- \omega\left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{108}$$

$$= \omega\left[\bar{r}_n - (1-G)(1-\rho)\hat{p}_n\right] + (\phi_{gt} - \rho\phi_{g,t-1}) + \rho d_{n,t-1} \tag{109}$$

$$- \omega G\Big(\bar{p}_{nt} - \rho\bar{p}_{n,t-1}\Big) + \omega\delta a_{nt} + \omega\Big[e_{nt} - G\big(\nu_{nt} - \rho\nu_{n,t-1}\big)\Big] \tag{110}$$

$$= \pi_{0,n} + \bar{\phi}_{gt} + \pi_1 d_{n,t-1} + \pi_2\bar{p}_{nt} + \pi_3\bar{p}_{n,t-1} + \pi_4 a_{nt} + \varepsilon_{nt} \tag{111}$$

Assume that the innovation in relationship capital e_{nt} is realized at the end of period t. Then e_{nt} will be uncorrelated with anything chosen at t or earlier, implying the combined error term ε_{nt} is uncorrelated with anything chosen at $t-1$ or earlier. The most efficient way to remove the reduced-form fixed-effect $\pi_{0,n}$ is to use forward orthogonal deviations. For any variable u, define the forward-orthogonal transformation

$$\ddot{u}_t = u_t - \frac{1}{T-t+1}\sum_t^T u_t$$

Applying this transformation to (111) yields

$$\ddot{d}_{nt} = \ddot{\bar{\phi}}_{gt} + \pi_1\ddot{d}_{n,t-1} + \pi_2\ddot{\bar{p}}_{nt} + \pi_3\ddot{\bar{p}}_{n,t-1} + \ddot{\pi}_4 a_{nt} + \ddot{\varepsilon}_{nt} \tag{112}$$

Estimate the transformed equation by removing the combined location-time fixed-effect $\tilde{\phi}_{gt} = \ddot{\phi}_{gt} - \rho\ddot{\phi}_{g,t-1}$ and instrumenting each transformed variable $\ddot{x}_t$ with the un-transformed x_t.[17] Under the assumptions, this procedure yields consistent estimates $\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3, \hat{\pi}_4$.

The next step is to use those estimates to construct an unbiased estimate of the combined fixed-effect. Let $\breve{x}_{nt}$ be the value of a variable x_{nt} after subtracting out the location-time mean. Then we can estimate marketing productivity (up to a fixed constant) as

$$\hat{\pi}_{0,n} = \frac{1}{T-2}\sum_{t=2}^{T-1}\Big[\breve{d}_{nt} - \hat{\pi}_1\breve{d}_{n,t-1} - \hat{\pi}_2\breve{\bar{p}}_{nt} - \hat{\pi}_3\breve{\bar{p}}_{n,t-1} - \hat{\pi}_4\breve{a}_{nt}\Big] \tag{113}$$

The final step is to isolate marketing productivity $\bar{r}_n$ by regressing the combined fixed-effect $\hat{\pi}_{0,n}$ on the prior of prices $\bar{p}_n$, which we proxy with the within-shop average price. The residual from this regression is proportional to marketing productivity.

[17] Instrumenting is necessary because $\ddot{d}_{n,t-1}$ is correlated with the error term by construction.

B.4 Two Markups

Suppose the set of goods J can be partitioned into a set of high markup goods $\bar{J}$ and low markup goods $\underline{J}$. Let us abuse notation and use the symbols $\bar{J}$ and $\underline{J}$ to also denote the number of goods in each set (much as J is also the total number of goods). We assume the average cost of goods across the two sets is identical:

$$\frac{1}{\bar{J}}\sum_{j\in\bar{J}}\kappa_j = \frac{1}{\underline{J}}\sum_{j\in\underline{J}}\kappa_j = \dot{\kappa}$$

The only difference is that the markup is μ on low cost goods and $a\mu$ on high cost goods (for some premium $a > 1$).

The first-order condition (5) now implies

$$x_j = \frac{\tilde{\alpha}_j DS}{J} + \frac{\sigma DS}{J} \times \begin{cases} \Phi^{-1}\left(1-\frac{\lambda}{\mu}\right) & \text{if } j \in \underline{J} \\ \Phi^{-1}\left(1-\frac{\lambda}{a\mu}\right) & \text{if } j \in \underline{J} \end{cases} \tag{114}$$

Define total spending on high- and low-markup goods as

$$\bar{X} = \sum_{j\in\bar{J}}\kappa_j x_j$$
$$\underline{X} = \sum_{j\in\underline{J}}\kappa_j x_j$$

which implies

$$X = \bar{X} + \underline{X} \tag{115}$$

Define average demand for each type of good as

$$\bar{\zeta} = \frac{1}{\bar{J}}\sum_{j\in\bar{J}}\kappa_j\tilde{\alpha}_j$$
$$\underline{\zeta} = \frac{1}{\underline{J}}\sum_{j\in\underline{J}}\kappa_j\tilde{\alpha}_j$$

Then

$$\begin{aligned}\bar{X} &= c\bar{\zeta}DS + c\dot{\kappa}\sigma DS\Phi^{-1}\left(1-\frac{\lambda}{a\mu}\right) \\ &= \frac{1}{Q}\left[c\bar{\zeta}D + c\dot{\kappa}\sigma D\Phi^{-1}\left(1-\frac{\lambda}{a\mu}\right)\right]\end{aligned}$$

$$\begin{aligned}\underline{X} &= (1-c)\bar{\zeta}DS + (1-c)\dot{\kappa}\sigma DS\Phi^{-1}\left(1-\frac{\lambda}{\mu}\right)\\ &= \frac{1}{Q}\left[(1-c)\bar{\zeta}D + (1-c)\dot{\kappa}\sigma D\Phi^{-1}\left(1-\frac{\lambda}{\mu}\right)\right]\end{aligned} \tag{116}$$

where $c = \bar{J}/J$

Define the surplus purchases of high- and low-markup goods as

$$\begin{aligned}\bar{\upsilon} &= \frac{\bar{X} - c\bar{\zeta}DS}{c\dot{\kappa}\sigma DS}\\ \underline{\upsilon} &= \frac{\underline{X} - (1-c)\bar{\zeta}DS}{(1-c)\dot{\kappa}\sigma DS}\end{aligned} \tag{117}$$

which implies the marginal product of inventory expenditure is

$$a\mu\left[1-\Phi(\bar{\upsilon})\right] = \lambda = \mu\left[1-\Phi(\underline{\upsilon})\right] \tag{118}$$

Define the overall aggregate surplus υ as in the main text. Then Equations 115, 116, and 117 imply

$$\upsilon = c\bar{\upsilon} + (1-c)\underline{\upsilon} \tag{119}$$

$$= c\Phi^{-1}\left(1-\frac{\lambda}{a\mu}\right) + (1-c)\Phi^{-1}\left(1-\frac{\lambda}{\mu}\right) \tag{120}$$

Using arguments parallel to those in Appendix A.1 one can show, for large $\bar{J}$ and $\underline{J}$, that

$$Y \approx \mathbb{E}[Y] = a\mu\left\{\bar{X} - \frac{c\sigma\dot{\kappa}D}{Q}\left[\Phi\left(\bar{\upsilon}\right)\bar{\upsilon} + \phi\left(\bar{\upsilon}\right)\right]\right\} + \mu\left\{\underline{X} - \frac{(1-c)\sigma\dot{\kappa}D}{Q}\left[\Phi\left(\underline{\upsilon}\right)\underline{\upsilon} + \phi\left(\underline{\upsilon}\right)\right]\right\} \tag{121}$$

$$= \bar{Y}(\bar{X}, Q) + \underline{Y}(\underline{X}, Q) \tag{122}$$

Define

$$\bar{G}(\bar{X}, Q) = \bar{X} - (1-\delta)\left(\bar{X} - \frac{\bar{Y}}{a\mu}\right) \tag{123}$$

$$\underline{G}(\underline{X}, Q) = \underline{X} - (1-\delta)\left(\underline{X} - \frac{\underline{Y}}{\mu}\right) \tag{124}$$

and observe that $G = \bar{G} + \underline{G}$ is the total additional inventory that must be sourced to maintain a total inventory of X. Then $G \leq ZK^{\beta}L^{\nu}$ is again the constraint on the firm's logistics. Then the first-order conditions of the Lagrangian (70) are

$$\vartheta\frac{\partial G}{\partial X} = \left\{\lambda - 1 + (1-\delta)\left[\Phi(\bar{\upsilon})\frac{\partial \bar{X}}{\partial X} + \Phi(\underline{\upsilon})\frac{\partial \underline{X}}{\partial X}\right]\right\}Q \tag{125}$$

$$\vartheta\frac{\partial G}{\partial Q}=\left\{\frac{X\lambda}{Q}-\frac{X}{Q}+(1-\delta)\left[\left(\frac{\bar{X}}{Q}+\frac{\partial\bar{X}}{\partial Q}\right)\Phi\left(\bar{\upsilon}\right)+\left(\frac{\underline{X}}{Q}+\frac{\partial\underline{X}}{\partial Q}\right)\Phi\left(\underline{\upsilon}\right)\right]\right\}Q-f \tag{126}$$

where Equation 125 follows because (118) implies

$$\frac{\partial Y}{\partial X}=\lambda$$

and (115) implies

$$\frac{\partial\bar{X}}{\partial X}+\frac{\partial\underline{X}}{\partial X}=1$$

and Equation 126 follows because (115) also implies

$$\frac{\partial\bar{X}}{\partial Q}+\frac{\partial\underline{X}}{\partial Q}=0$$

Taking the partial derivatives of Equation 116 with respect to X and Q yields

$$\begin{aligned}\frac{\partial\bar{X}}{\partial X}&=\frac{\sigma Dc\dot{\kappa}}{Q}\frac{1}{\phi(\bar{\upsilon})}\frac{-\partial\lambda/\partial X}{a\mu}\\ \frac{\partial\bar{X}}{\partial Q}&=-\frac{\bar{X}}{Q}+\frac{\sigma Dc\dot{\kappa}}{Q}\frac{1}{\phi(\bar{\upsilon})}\frac{-\partial\lambda/\partial Q}{a\mu}\end{aligned} \tag{127}$$

with parallel expressions for the partial derivatives of $\underline{X}$. Apply the implicit function theorem to (120) yields

$$\begin{aligned}\frac{\partial\lambda}{\partial X}&=-\frac{\frac{Q}{\dot{\kappa}\sigma D}}{\frac{c}{a\mu\phi(\bar{\upsilon})}+\frac{1-c}{\mu\phi(\underline{\upsilon})}}\\ \frac{\partial\lambda}{\partial Q}&=-\frac{\frac{X}{\dot{\kappa}\sigma D}}{\frac{c}{a\mu\phi(\bar{\upsilon})}+\frac{1-c}{\mu\phi(\underline{\upsilon})}}\end{aligned} \tag{128}$$

Sub (128) into (127) and rearrange to show that

$$\frac{\bar{X}}{Q}+\frac{\partial\bar{X}}{\partial Q}=\frac{\partial\bar{X}}{\partial X}$$

with a similar expression for $\underline{X}$. Then the first order conditions (125) and (126) simplify to

$$\vartheta\frac{\partial G}{\partial X}=\left\{\lambda-1+(1-\delta)\left[\frac{\partial\bar{X}}{\partial X}\Phi(\bar{\upsilon})+\frac{\partial\underline{X}}{\partial X}\Phi(\underline{\upsilon})\right]\right\}Q \tag{129}$$

$$\vartheta\frac{\partial G}{\partial Q}=\left\{\lambda-1+(1-\delta)\left[\frac{\partial\bar{X}}{\partial X}\Phi\left(\bar{\upsilon}\right)+\frac{\partial\underline{X}}{\partial X}\Phi\left(\underline{\upsilon}\right)\right]\right\}X-f \tag{130}$$

From these equations it is clear that if Q and f are positive then ϑ cannot equal zero, implying the constraint binds and Proposition 2 still holds.

Equation 120 implies that

$$\lambda = \boldsymbol{\lambda}(\upsilon, c, a, \mu)$$

for some function $\boldsymbol{\lambda}$. Then

$$\bar{\upsilon} = \Phi^{-1}\left(1 - \frac{\boldsymbol{\lambda}(\upsilon, c, a, \mu)}{a\mu}\right) \tag{131}$$

$$\underline{\upsilon} = \Phi^{-1}\left(1 - \frac{\boldsymbol{\lambda}(\upsilon, c, a, \mu)}{\mu}\right) \tag{132}$$

are both functions of only (υ, c, a, μ), implying

$$\frac{\partial \bar{X}}{\partial X} = \frac{\frac{c}{a\mu\phi(\bar{\upsilon})}}{\frac{c}{a\mu\phi(\bar{\upsilon})} + \frac{1-c}{\mu\phi(\underline{\upsilon})}} \tag{133}$$

$$\frac{\partial \underline{X}}{\partial X} = \frac{\frac{1-c}{\mu\phi(\underline{\upsilon})}}{\frac{c}{a\mu\phi(\bar{\upsilon})} + \frac{1-c}{\mu\phi(\underline{\upsilon})}} \tag{134}$$

are as well. We can thus define

$$\aleph(\upsilon, c, a, \mu) = \frac{\partial \bar{X}}{\partial X}\Phi\left(\bar{\upsilon}\right) + \frac{\partial \underline{X}}{\partial X}\Phi\left(\underline{\upsilon}\right)$$

Divide (130) by (129) and simplify to yield

$$\frac{f}{\boldsymbol{\lambda}(\upsilon, c, a, \mu) - 1 + (1-\delta)\aleph(\upsilon, c, a, \mu)} = \frac{B}{1 - (1-\delta)\aleph(\upsilon, c, a, \mu)} \tag{135}$$

Define $\dot{X} = a\mu\bar{X} + \mu\underline{X}$ and assume, as in the main text, that δ is the same across all shops. Then the solution $\boldsymbol{\upsilon}(c, a, \mu, B, \delta, f)$ can be subbed into (121) and rearranged to yield

$$\log\left[\frac{Q\dot{X} - QY}{Q\dot{X}}\right] = -\log\mathcal{I} + h(\log c, \log a, \log\mu, \log(f/B), \log D, \log Q, \log\dot{X})$$

where the left-hand side is the log of unsold inventory valued a *retail prices*. This is in contrast to the version of the estimator in the main text, which values unsold inventory at cost. This equation can be used to construct a more general version of the estimator in Section 5.3. Aside from changing the dependent variable, the estimator must expand the terms in the control function to include the share of high-markup goods and the premium on that markup, and to control for (target) inventory valued at the retail price.

Though this appendix assumes there are only two markups, the argument does not hinge on the number and we conjecture it could be generalized. The limitation is likely to be practical rather than theoretical. It may not be feasible to measure a whole set of markups $\{a\}$ and their distribution $\{c\}$. Even if measurement is feasible, the curse of dimensionality implies adding

more terms to the nonparametric estimator will require a possibly infeasible number of observations. But the two-markup case does capture the classic case of a store that prices some goods as "loss leaders" that earn little profit but attract customers in the hopes that they also buy higher profit complementary goods.

C Data Appendix

C.1 Definitions of Variables Used to Estimate Productivity

Sales Y_{nt}: For every day during the response week where the shop was open, respondents were asked "How much did you earn from goods sold on [day]?"

Cost of goods sold Y_{nt}/μ_{nt}: For every day during the response week where the shop was open, respondents were asked "How much would it cost to replace all goods sold on [day]? Consider only the price that you pay for the goods."

Markup μ_{nt}: We calculate the average daily ratio of Sales to Cost of goods sold.

Number of customers D_{nt}: For every day during the response week where the shop was open, respondents were asked "How many customers entered the store or stopped to look at your products on [day]? Include customers picking up phone/online orders."

Frequency of restocking Q_{nt}: For every day during the response week (unless the shop was closed all week) the respondent was asked "Did you buy any new inventory (goods for resale) on [day]?". We calculate the average of this dummy across all days that the shop was contacted, which is the probability of restocking, and define a set of dummies for the values.[18]

Inventory purchased B_{nt}: If the respondent did buy inventory, they were asked "How much did you spend on goods for resale on [day]? Enter in ZMW." We calculate the average expenditure per restocking event.

Target Inventory after restocking X_{nt} Immediately after the response week (the Sunday survey, which was done on Monday of the following week) the respondent was asked "Think about the time over the past week when your stock of inventory was at its highest level (you had more goods for sale than at any other point). If you were to buy those goods today, what would be the total cost of the goods?" We take this "highest level" to be the post-restocking target size of the inventory.

[18] To adjust for possible non-response during a few days of the response week, in practice we calculate the fraction of surveys where the respondent said they had bought goods for resale. The dummies are defined to indicate each possible value for this fraction (there are roughly 10 values in practice).

Total inventory available for purchase throughout the week $Q_{nt}X_{nt}$ During the Monday survey of each week (done on Tuesday of the reference week) the respondent was asked "Consider all the inventory that you had at the beginning of the day on Monday morning. If you purchased all these goods today, how much would you have to pay?" We add this to the product of B_{nt} and the number of restocking events.[19]

Fixed cost of restocking f_{nt}**:** If the respondent did buy inventory, they were asked three questions about costs of restocking: "How much did you have to pay to have goods delivered to the shop on [day]?" ; "What was the cost of transport for you/workers to bring goods for resale back to the shop?" ; and "How many hours on [day] did you and your workers spend IN TOTAL visiting markets/suppliers, picking up goods for resale, and stocking shelves?" The latter was converted to a monetary value using the hourly wage that the respondent estimates they would have to pay to hire a casual worker. If the respondent was unable to answer the question, we imputed the average response among all respondents in the market. There was one market with relatively few respondents (Kalingalinga) where no one was able to answer the question. For respondents in that market we assign the mean from a neighboring market (Mutendere). We calculate f_{nt} as the average across all restocking events.

Value of goods lost to depreciation $\delta_{nt}(X_{nt} - Y_{nt}/\mu_{nt})$**:** Immediately after the response week (the Sunday survey, which was done on Monday of the following week) the respondent was asked "Over the past week, were there any goods that you threw away, gave away, or kept for yourself because they had spoiled or were about to spoil?" If they answered yes, they were asked "What would it cost to replace these goods? Consider only the order price of goods." If they answered no to the first question, we infer that the value is zero.

Investment: During the Tuesday survey of each round of the panel survey (asked on Wednesday of the response week), the respondent was asked "Since [prior survey date], have you bought any new equipment for this business?" If the answer was yes, they were asked "What was the total cost of the new equipment? If the equipment was received as a gift, enter your best guess of what it would cost to buy it yourself." They were then asked "Since [prior survey date], have you sold or discarded any of this business's equipment?" If the answer was yes, they were asked "What would the equipment sold or discarded or given away cost if bought today?" Net investment is defined as the difference between investment and disinvestment.

Capital K_{nt}**:** During the baseline survey the respondent was asked for the resale value of all assets related to the business (not including the structure). We calculate the value of capital

[19] To adjust for possible non-response during a few days of the response week, in practice we calculate the fraction of surveys where the respondent said they had bought goods for resale, then multiply that by 7.

at baseline as the sum of the values of these assets (excluding the value of signs, which are assumed to be advertising investment). For each round we started with the value of assets in the previous round (or the baseline, in the case of Round 1), and added the value of investment for that round. Since the response rate for Round 1 was low, in Round 2 we asked about investment since baseline and added that value to the baseline value (to avoid missing investment made during Round 1).

Labor L_{nt}: For every day during the response week (unless the shop was closed all week) the respondent was asked three questions: "How many hours did YOU work on on [day]?" ; "How many hours did PAID WORKERS work on on [day]?" ; "How many hours did UNPAID WORKERS work on on [day]?" The sum of these three is taken as labor for that day. We calculate the average for all days that the respondent answered the survey.

Hours H_{nt}: Every day during the response week (unless the shop was closed all week) the respondent was asked "How many hours was the shop open to customers on [day]?"

Value of unsold inventory W_{nt}: Calculating the actual difference between inventory available and inventory sold is extremely noisy. Instead we ask each respondent during the Sunday survey (which was done on Monday of week after the response week) "Consider all the inventory that you had at the end of the day on Sunday evening. If you purchased all these goods today, how much would you have to pay?" This is the most accurate measure of the residual inventory at the end of the week.

Time-varying Advertising Investment a_{nt}: During the Tuesday survey of each round of the panel survey (asked on Wednesday of the response week), the respondent was asked "Since [prior survey date], have you spent any money on signs, fliers, social media ads, radio, text messages, etc., for attracting customers? This can include showing information about your business, the goods you sell, or the prices you charge" If the answer was yes, they were asked "How much did you spend?" The respondent was also asked "Since [prior survey date], how many hours in total did you/your employees spent advertising/attracting customers (calling/texting customers, handing out flyers in the market, etc.)" This latter response was converted to a monetary value using the same inferred wage as used for the cost of restocking above. We sum these three values.

Fixed Advertising Investment $\bar{a}_n$: During the baseline survey the respondent was asked for the resale value of all assets related to the business (not including the structure). One category of assets, signs used for advertising, is excluded from the calculation of capital and instead used as the value of fixed advertising expenditure. Since this is zero for many retailers we use the inverse hyperbolic sine when estimating equation 40.

Price level P_{nt}: During the baseline survey the respondent was asked to name their top three goods by sales, and the unit of purchase and sale. During subsequent rounds we asked about the prices and order prices of each of these goods, as valued using the unit of purchase or sale, regardless of whether they were still the top three goods (comparable to a price basket used to calculate a standard price index). If the respondent no longer purchased or sold in the unit, they were asked for the new unit, and that unit was used in subsequent rounds (until and unless the unit changed again). Respondents were asked to estimate the weight or volume of the sales unit and the conversion rate between buying and selling units at baseline, Round 4, and Round 7. These weights were used to calculate a homogenized (per gram/ml) price. These homogenized prices were used to calculate a weighted average price, where the weights were the value of sales of that good relative to the full set of three goods.