Working paper ZMB-22099 September 2024

Measuring and estimating retail productivity

Brenda Samaniego de la Parra
Ajay Shenoy

SGB EVIDENCE FUND
argidius
ASPEN NETWORK OF DEVELOPMENT ENTREPRENEURS
aspen institute
IGC
DIRECTED BY
LSE
UNIVERSITY OF OXFORD

Measuring and Estimating Retail Productivity

Brenda Samaniego de la Parra*
Ajay Shenoy†

November 19, 2024

Abstract

We develop and apply a new method to estimate the productivity of small and medium retailers in developing countries. The method assumes a successful retail shop must 1) attract customers, 2) source and stock inventory, and 3) choose the right mix of products from the right suppliers. Our method estimates the productivity of a shop across all three dimensions. We apply our method to a novel retail-specific dataset collected through high frequency surveys of a sample of independent retailers in Lusaka, Zambia. Our results suggest the dimensions of productivity are correlated but also contain substantial independent variation. An entrepreneur can be highly productive in one aspect of running a shop but unproductive in others. Under these conditions, our model predicts that standard estimators would yield deeply misleading results. Policy interventions based on those estimators could actually lower social welfare.

JEL Codes: D24, L23, L81, C13

*University of California, Santa Cruz

†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.

1 Introduction

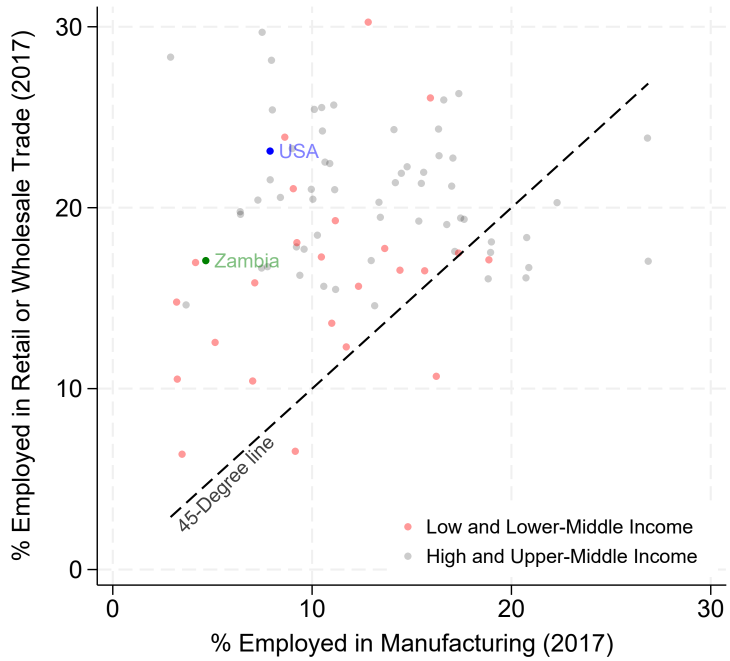
The retail sector employs more people than manufacturing in nearly every country at every stage of development (see Figure 1).[1] Productivity in the retail sector determines the incomes of the hundreds of millions of workers it employs. It is also crucial for the many billions more who rely on the retail 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.

But measuring retail productivity raises novel challenges. A retailer does not "produce" a physical good but a match between a consumer and a good produced by others. The unit of output, a sale, literally does not exist separately from consumer demand. This key feature is missing from standard models of production designed for manufacturing and agriculture, where inputs are transformed into physical output that can be measured regardless of whether it is sold. Modeling it effectively requires taking seriously the idea that productivity may have multiple dimensions that span distinct stages of production. Ignoring the distinction can yield biased estimates or misleading conclusions. Policy recommendations based on standard methods may be invalid and even welfare-reducing.

This paper proposes a method to estimate retail productivity across three dimensions. We assume that a productive retail shop 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 productivity, 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 productivity." Second, the shop's labor, capital, and "logistical productivity" 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 productivity."

[1] The figure combines retail and wholesale trade because we are unaware of any internationally comparable dataset that measure them separately. Among countries that do 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).

The three-stage model makes novel predictions that run counter to what would be derived from a standard one-stage model of productivity. First, it predicts that the marginal product of additional inventory may be higher *or lower* for shops with higher inventory choice productivity. 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. A productive shop will on average suffer less costly stock-outs, lowering the returns to the buffer. This is in stark contrast to standard models, where higher productivity always implies a higher marginal product.

Second, the model predicts that, all-else-equal, a shop with higher logistical productivity 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 is predicted by the standard model, 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 productivity.

Finally, the model predicts that an increase in logistical productivity alone may actually lower social welfare. This is analogous to the thought experiment of the grocer who increases stock without selling everything. The shop will be able to increase its inventory, but unless it is also productive in marketing and inventory choice, 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 higher productivity simply increases output. As we explain below, these differences have critical implications for how standard productivity estimators may be (mis)interpreted, and how policy recommendations based on these estimates may be invalid.

The model implies that each of the three dimensions of productivity can be estimated with simple fixed-effects or instrumental variables regressions. To estimate these equations we construct a unique panel dataset from a 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 geocoded screener and a detailed baseline survey of assets, suppliers, and managerial practices. The sample includes shops at every level 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 show that many of the basic premises of the model hold in our data. There is widespread variation in profitability. There is large dispersion in the prices paid by shops for even homogeneous goods for resale. Shops make different choices of inventory—even choosing different varieties of onions—and those varieties earn different average profits. Shops vary drastically in the number and distance of their suppliers, suggesting differences in logistical sophistication. There are big differences in the number of customers attracted, and whether they use even basic means to attract customers (such as posting printed signs).

Our estimates of marketing, logistical, and inventory choice productivity also behave as would be expected. All three measures are well-correlated with profitability, both in- and out-of-sample. Estimates constructed from only the first seven rounds of data are better predictors of gross profit (value-added) in the final round than standard measures of managerial practices. We also show that, as the model predicts will hold under certain conditions, there is a *negative* correlation between logistical productivity and capital.

We then show that our measures add value in understanding productivity and designing optimal policies. We show that our measures jointly explain nearly 3 times as much of the variation in future value-added than the most comparable standard estimator of univariate Hicks-neutral productivity (one that assumes productivity is fixed within-firm). That is in part because the three dimensions of productivity, though correlated, have substantial independent variation.

This last fact has critical implications for research and policy about improvements to productivity. An analysis of whether the economy has efficiently allocated capital could find large misallocation even though capital markets are efficient. Shops with high logistical productivity but low inventory choice productivity would optimally choose lower levels of capital, but standard estimates of productivity would imply that they should have higher levels of capital. A researcher using standard estimates would incorrectly infer that these shops are prevented from getting enough capital by some costly market friction. The consequences of using standard estimates are potentially even graver for policy. An intervention to provide inventory grants, if targeted to shops measured as more productive under standard estimators, could have disappointing or even harmful impacts. Firms with high logistical productivity but low inventory choice productivity would buy large quantities of poorly chosen inventory that is ultimately discarded.

1.1 Relation and Contribution to Literature

This paper joins an old literature, dating back to Solow (1957), seeking to measure and interpret productivity and its role in economic development. Though too vast to describe in broad strokes, much of the literature does have one feature in common: a focus on manufacturing or agriculture. Lagakos (2016), the most notable exception, proposes a theory for why *aggregate* retail productivity in developed countries is higher than in developing countries (which You, 2021, tests in turn-of-the-century Boston). Foster et al. (2006) also studies aggregate retail productivity, applying decomposition methods disentangle how much of the growth in output-per-worker in the U.S. retail sector comes from higher average (firm-level) output versus different forms of reallocation. Our study is novel in defining and measuring *firm-level* productivity that is grounded in a model of retail production. Our measures push the frontier of this literature beyond the univariate measures that are the norm by joining the as-yet small literature that models production within the firm as a multi-part process.[2]

In each stage of our model, productivity is in large part a reflection of how well that stage is managed. This feature of the model ties it to the literature on management practices. Prior work measures the quality of management using binary or categorical measures of whether the firm employs each of several practices (Bloom et al., 2012; McKenzie and Woodruff, 2017; Dalton et al., 2021). These survey-based measures are inexpensive and have been collected across many countries. Though these measures in aggregate are often correlated with positive outcomes, the impact on productivity of any individual management practice is less obvious and may vary across countries and industries. Our contribution is to derive measures of productivity from a model of retail production and management. This approach ensures the measures reflect realized outcomes (and thus actual success) while also being directly linked to a specific dimension of management. Inventory choice productivity, for example, is a direct measure of the firm's ability to select products currently in demand. As we show, these measures are stronger predictors of profitability and value-added than survey questions about management practices. We also find that our measures are not well-correlated with the most common "standard" management practices, suggesting the standard may need to be adjusted for surveys of retail firms.

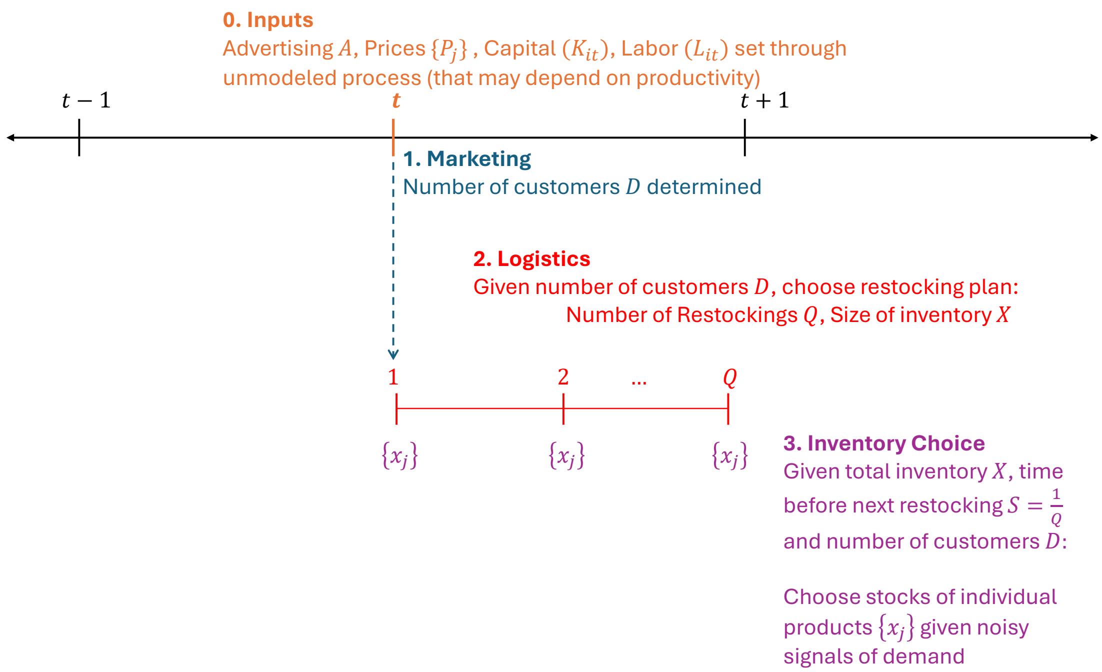
Our project also contributes to the literature on production function estimation. The modern approach to production function estimation was designed to address what is sometimes called "transmission bias" (Ackerberg et al., 2007), the fact that a firm's choice of inputs is partly a function of unobserved productivity. There are broadly two approaches to handle this problem: "structural" or "choice-based" methods, and dynamic panel methods. What all of these methods have in common is that they were designed for and originally applied to the manufacturing sector. Olley and Pakes (1996) studied telecommunications manufacturers in the U.S., Blundell and Bond (2000) a sample of U.S. manufacturers, Levinsohn and Petrin (2003) and Ackerberg et al. (2006) a set of manufacturing industries in Chile, and Gandhi et al. (2017) those same Chilean industries plus the same industries in Colombia.[3] Perhaps as a result, all these methods define production as the conversion of physical inputs into output. To be precise, they assume there is a univariate dimension of productivity, and that output depends only on productivity and a set of productive inputs (generally capital and labor). As we show formally in Section 3.4, applying these most commonly used methods to a model of retail production like the one we propose will yield measures of productivity that are either biased or prone to misinterpretation.

Finally, our work is related to a literature in industrial organization that studies the location

[2]A few studies, e.g. Kremer (1993), model multiple stages but with uniformity across the stages, and not with the goal of measuring productivity.

[3]Wooldridge (2009), which does not present an empirical application, is ultimately a more efficient version of Levinsohn and Petrin (2003). Ackerberg et al. (2015), the published version of Ackerberg et al. (2006), presents the same method as the working paper but does not include the empirical application.

Figure 2: Model Timing



and pricing decisions of retailers (Varian, 1980; Ehrlich and Fisher, 1982; Bagwell and Ramey, 1994; Faig and Jerez, 2005; Armstrong and Zhou, 2011; Johnson, 2017).[4] While much of the earliest literature is theoretical, more recent work has proposed and estimated structural models of how retailers acquire customers (for example Smith, 2004; Thomassen et al., 2017; Ellickson et al., 2020). Our project has an entirely different objective: to produce a flexible method for estimating productivity in retail. As a result, we propose a model that captures logistics and inventory choice as well as customer acquisition. Our model implies a simple set of estimators that can be taken to the data.

2 Model

The design of any model is a balance of trade-offs, and ours prioritizes making the minimal assumptions needed to derive three intuitive estimating equations. This balance reflects its purpose: it is a platform for defining and estimating productivity, 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 to infer productiv-

[4]A few studies, like Blair and Lewis (1994), study the principle-agent problem between retailers and their suppliers.

ity. This is analogous to how the production function literature does not model the choices of capital and labor beyond imposing the minimal restrictions on those choices necessary to nonparametrically control for them.[5] These unmodeled 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 productivity." 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 productivity." 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. Productive shops 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 productivity."

The timing is as shown in Figure 2. In each period:

0. Shop locations, prices, costs, advertising expenditures, capital, and labor are determined through some unmodeled process. These quantities may be correlated with productivity
1. The firm makes all choices related to marketing (prices and advertising) which determines 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. At each restocking the shop draws a noisy signals of demand for each product and 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.

2.1 Inventory Choice

There are J goods that the shop could potentially stock. At each restocking the shop receives 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 interval, known with certainty, and pre-determined by the choices made in the marketing part of the model (see below).

[5] See the review in Ackerberg et al. (2007) for examples.

Each customer buys the same quantity α_j/J, which is fixed within the period but unknown at the time the 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 2.3). But Appendix B.1 shows that allowing for an intensive margin is straightforward and does not change the final estimating equation.

Every time the shop draws a new α_j it 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, making it more productive. 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.[6] Then the 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 a period 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 over the period 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 firm 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

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

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

To make the problem tractable, we assume that each shop has a fixed markup

$$P_j = \mu \kappa_j$$

that may vary across shops but is the same across goods.

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

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

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

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

This expression shows that one common intuition from manufacturing—that holding fixed inputs, a more productive firm has higher marginal product—does not apply to retail. A firm 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 very much. 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

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

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, it is shown (in Appendix A.1) that

Proposition 1

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

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 $\bar{\kappa}$ and noisier signals (higher σ) have lower sales. We define the level of inventory choice productivity as

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

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 productivity. 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.

2.2 Logistics

2.21. 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 resocking 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 productivity. 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

$$Z K^\beta L^\nu \tag{14}$$

though it is easy to generalize the derivation and estimation.

2.22. 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 $\mathcal{D}(S)$ of its value, 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 productivity:

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 standard production function estimated in manufacturing. But as we explain below, the difference is crucial in how we interpret the impact of an all-else-equal increase in logistic productivity. 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 productivity:

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 6.3 how Equation 17 can be combined with Equation 10 to derive an estimating equation.

2.3 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 productivity 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 expenditures (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 productivity, 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 productivity (or their absence). If not maintained, relationship capital depreciates at a rate $(1 - \rho)$. The law of motion for relationship capital is

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

2.4 Profit and Productivity

As in the standard model, more productive firms are more profitable. But in this model, each dimension of productivity raises profit for a different reason.

Define the adjusted gross profit over some period 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:

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

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 productivity to adjusted profit. The expression in braces is profit per customer, which is scaled by the number of customers (and marketing productivity). As long as profit-per-customer is positive, an increase in marketing productivity will proportionally increase profits. Per-unit profit depends on the other forms of productivity. An increase in inventory choice productivity reduces the negative term inside the braces, which represents the value of goods unsold. That is, a firm with higher inventory choice productivity earns higher profit because it throws away less of its stock. Finally, higher logistical productivity 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.[7]

3 Differences with the Standard Model

The familiar model of production, which models $Y = Zf(K, L)$ and has been commonly applied to manufacturing and farming, carries several predictions that have become the basis for many policies aimed at increasing production. Our model of retail differs in key ways that alter or even reverse these policy prescriptions.

3.1 The Marginal Product of Inventory is Not Necessarily Higher for More Productive Firms

The standard model predicts that the marginal product of an input—say, capital—is strictly increasing in productivity:

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

which is strictly positive under the standard assumption that the production function al-

[7]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 $\mathcal{M}(Z) > 0$.

ways has positive returns in all inputs. Given two firms with equal levels of capital and labor, a program that hands out extra capital or labor should be targeted at the more productive firm. Conversely, an efficient market would allocate more inputs to the more productive firm.

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

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

This term is positive if and only if $X > \zeta DS$. An increase in inventory productivity—that is, a *decrease* in σ or $\bar{\kappa}$—raises the marginal product of inventory if and only if $X < \zeta DS$. In other words, it is only efficient to give the more productive shop additional inventory if both shops have less inventory than needed to satisfy expected demand.

The result is unconventional but not unintuitive. 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 one that makes better predictions can use the marginal dollar of inventory spending more efficiently because it can predict and counter the costliest stockouts.

3.2 A Shop with Higher Productivity May Choose Less Capital

Our model has an even starker contrast with the standard model in the case of capital and labor. Equation 31 implies that in the standard model a firm with higher productivity 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 productivity 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 productivity (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.

3.3 Raising Only One Dimension of Productivity 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 productivity 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 productivity will be portioned between increasing actual sales and increasing the quantity of goods spoiled or lost. Put another way, one "dollar" of extra productivity 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 productivity) and the extent to which demand for individual products is matched to the actual stocks (which depends on inventory choice productivity). 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 productivity always raises output and sales and is always welfare-enhancing. The contrast highlights the danger of both using a manufacturing-centric model to study retail, and of treating productivity as uni-dimensional. Making a firm more productive in one dimension without improving its productivity in other dimensions would at best attenuate the impact, and at worst cause more harm than good.

3.4 Consequences for Standard Productivity Estimators

Applying standard productivity estimators to retail or any sector with elements featured in our model could yield misleading conclusions. The estimators may be biased, and even if not the differences with the standard model imply they cannot be interpreted as simple univariate measures of total factor productivity.

Endogeneity and Biased Inference: Standard methods of production function estimation (e.g. Olley and Pakes, 1996; Blundell and Bond, 2000; Levinsohn and Petrin, 2003, and subsequent papers following either approach) put sales on the left-hand and capital and labor on the right-hand side of an estimating equation. The specific procedure and the moment conditions vary across methods, but this basic equation is the same. The goal is to disentangle a (possibly nonparametric) function of capital and labor, assumed common across firms, from Hicks-neutral productivity, assumed to vary across firms and possibly across time within firm.

Standard methods will estimate an equation of this form:

$$\log Y_{nt} = \omega_{nt} + \beta \log K_{nt} + \nu \log L_{nt} + \varepsilon_{nt} \tag{34}$$

where ω_{nt} is Hicks-neutral productivity and ε_{nt} is an independent error term (usually interpreted as measurement error). Different assumptions about the stochastic process of ω_{nt} imply different moment conditions.

Superficially this regression resembles Equation 36, which can be rearranged as

$$\log Y_{nt} = \log Z_n + \beta \log K_{nt} + \nu \log L_{nt} + \varepsilon_{nt} - \log\left[\delta(X_{nti} - Y_{nt}/\mu)\right] \tag{35}$$

where we have allowed for measurement error in sales.

Comparing (34) and (35) shows that a standard estimator does not account for the value of unsold inventory, which is absorbed into the error term. If it is correlated with capital or labor, it may bias the estimates. The size of that bias depends on the average quantity of lost or wasted inventory, and whether it is correlated with other factors of production.

Incorrect Counterfactuals: There is an even bigger risk that the estimate, even if unbiased, will be misinterpreted. If the correlation between the value of wasted inventory and the factors of production is small, then the estimate of ω_{nt} will be close to an unbiased estimate of logistical productivity $\log Z_n$. But an econometrician using standard methods would interpret it as overall productivity, and calculate counterfactuals accordingly. These calculations may yield misleading conclusions.

For example, a standard calculation would imply that, among two firms with similar levels of capital, the one with higher estimated productivity has a higher marginal product of capital. But Proposition 4 implies that the opposite may hold if logistical productivity is not strongly correlated with the other two dimensions of productivity. Giving capital to the shop that appears to have higher productivity would yield less output than giving the same capital to the other shop.

For the same reason a measure of misallocation in the spirit of Hsieh and Klenow (2009) might imply large misallocation because the firms that appear most productive do not have the most capital. One might conclude that there are serious flaws in rental or financial markets that keep productive entrepreneurs from getting enough capital. But the same proposition implies the negative relationship could hold in an efficient market because retailers with efficient logistics but weak marketing or inventory choice would rationally choose to hold less capital, and this choice would be socially efficient.

Flawed Policy Recommendations and Welfare Calculations: A policy recommendation is a prediction of which government intervention will produce the best counterfactual. If the counterfactuals are flawed, the recommendation will be as well. Continuing the earlier example, suppose a government wanted to distribute a fixed budget of equipment with the goal of maximizing output. It would need to target shops with the highest marginal product of capital. But a targeting policy based on standard estimates of productivity—which is logistical productivity alone—may target the wrong firms.

Worse still, it may actually reduce social welfare. Section 3.3 shows that giving capital to shops that are unproductive in marketing and inventory choice may lower welfare by increasing waste in the economy (if there are negative externalities that come with garbage or spoiled food). The same holds even for an intervention—say, a business training session—that focuses on logistics. Raising logistical productivity without also raising the other dimensions of productivity would run a similar risk of increasing the total inventory that is wasted.

4 Estimation

Assume there is a panel that records observations of each shop $n = 1, \dots, N$ during each month $t = 1, \dots, T$. We show how to use the results of Section 2 to estimate each of the three dimensions of productivity. We explain how each quantity in the estimating equations is measured in Appendix C.

4.1 Estimating Logistical Productivity

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

$$\log \tilde{Y}_{nt} = \log Z_n + \beta \log K_{nt} + \nu \log L_{nt} + \log Q_{nt} \tag{36}$$

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 productivity Z_n is fixed within shop over time, 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 productivity can be estimated as

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

4.2 Estimating Inventory Choice Productivity

Equation 10 shows how we can construct an estimator for inventory choice productivity. 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 available for sale during the week:

$$\frac{Q_{nt}X_{nt} - Q_{nt}\frac{Y_{nt}}{\mu_{nt}}}{Q_{nt}X_{nt}} = \frac{D}{\mathcal{I}}\frac{\Phi(\upsilon)\upsilon + \phi(\upsilon)}{Q_{nt}X_{nt}} \tag{37}$$

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

$$\log W_{nt} = -\log \mathcal{I}_n + h(\log X_{nt}, \log D_{nt}, \log \mu_{nt}, \log \frac{f_{nt}}{B_{nt}}, Q_{nt}) \tag{38}$$

where $h(\cdot)$ is a nonparametric function. Equation 38 can be estimated using a non-parametric fixed-effects regression, and inventory choice productivity would be the following function of the estimates:

$$\log \hat{\mathcal{I}}_n = -\frac{1}{T}\sum_{t=1}^{T}\Big(\log W_{nt} - \hat{h}(\log X_{nt}, \log D_{nt}, \log \mu_{nt}, f_{nt}, Q_{nt}, \log B_{nt})\Big) \tag{39}$$

4.3 Estimating Marketing Productivity

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 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

$$d_{nt} = \omega r_{nt} - \omega \left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{40}$$
$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + e_{nt}\Big) - \omega \left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{41}$$

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

$$d_{nt} = \omega r_{nt} - \omega \left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{42}$$
$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + \delta_2 \tilde{a}_{nt} + e_{nt}\Big) - \omega \left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{43}$$
$$= \omega\Big(\bar{r}_n + \delta_1 \bar{a}_n + (1-G)\hat{p}_n\Big) - \omega \left[G\bar{p}_{nt} + \delta_2 \tilde{a}_{nt}\right] + \phi_{gt} + (G\nu_{nt} + e_{nt}) \tag{44}$$

Equation 44 is the structural form of the estimating 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:

$$d_{nt} = \pi_{0,n} + \bar{\phi}_{gt} + \pi_2 \bar{p}_{nt} + \pi_4 \tilde{a}_{nt} + \varepsilon_{nt} \tag{45}$$

The second step is to isolate marketing productivity $\bar{r}_n$ by estimating the expression in the first parenthetical expression in Equation 44 with the combined fixed-effect $\hat{\pi}_{0,n}$ on the left-hand side:

$$\hat{\pi}_{0,n} = \pi_0' + \pi_1' \bar{a}_n + \pi_2' \hat{p}_n + \xi_n \tag{46}$$

The residual from this regression

$$\hat{\bar{r}}_n = \hat{\pi}_0 - \hat{\pi}_1' \bar{a}_n - \hat{\pi}_2' \hat{p}_n \tag{47}$$

is is some combination of marketing productivity and estimation error from the first stage, and thus an unbiased estimator for marketing productivity. 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 45 and 46 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.

5 Data and Context

5.1 Description of the Lusaka Retail Productivity Study

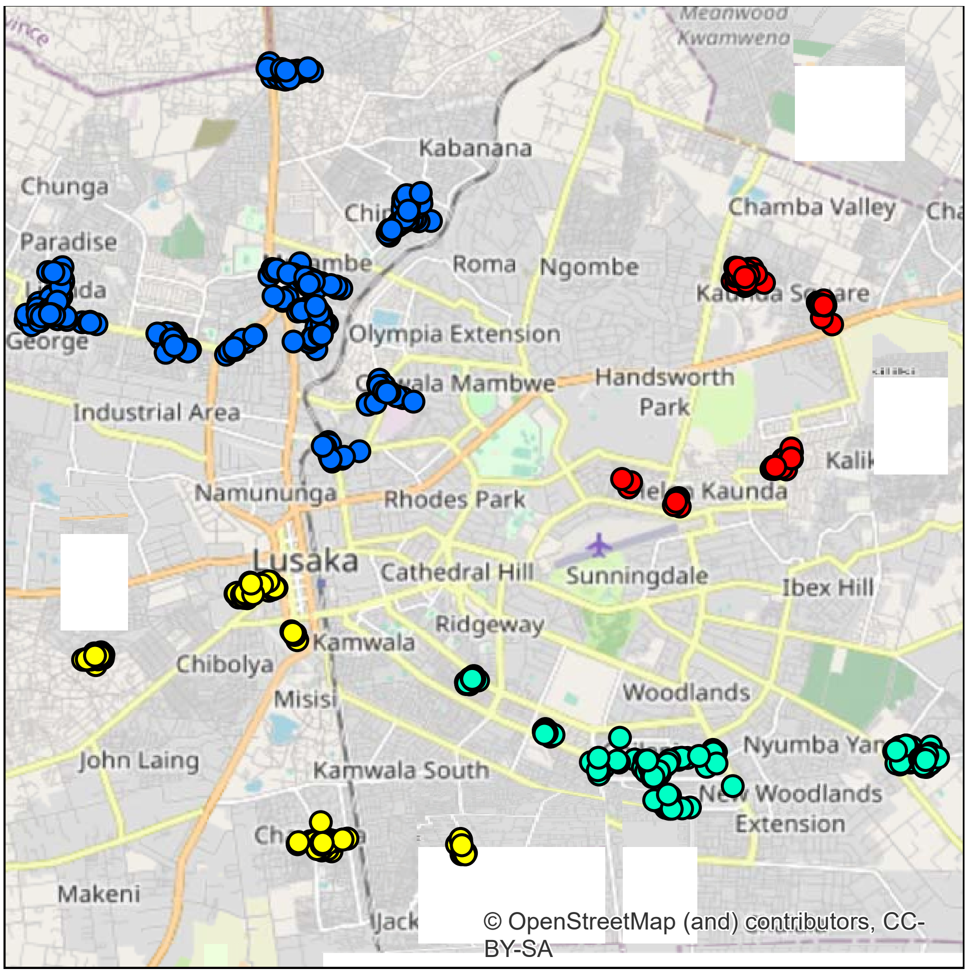
The biggest challenge to studying retail production is the lack of comprehensive data, especially in a developing country. Though there are many surveys and (in more developed countries) administrative datasets, these capture basic accounting concepts like sales and wages. Many key aspects of the retail operation, like the number of customers or the replacement value of goods sold, are not recorded in any administrative record or asked in any existing survey. The challenge is compounded in a developing country, where many managers do not keep written records of even standard accounting concepts.

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

Our study focused on 25 markets in Lusaka, Zambia. These markets were chosen 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 3000 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 1000 shops. Figure 3 shows the locations of the sample shops.

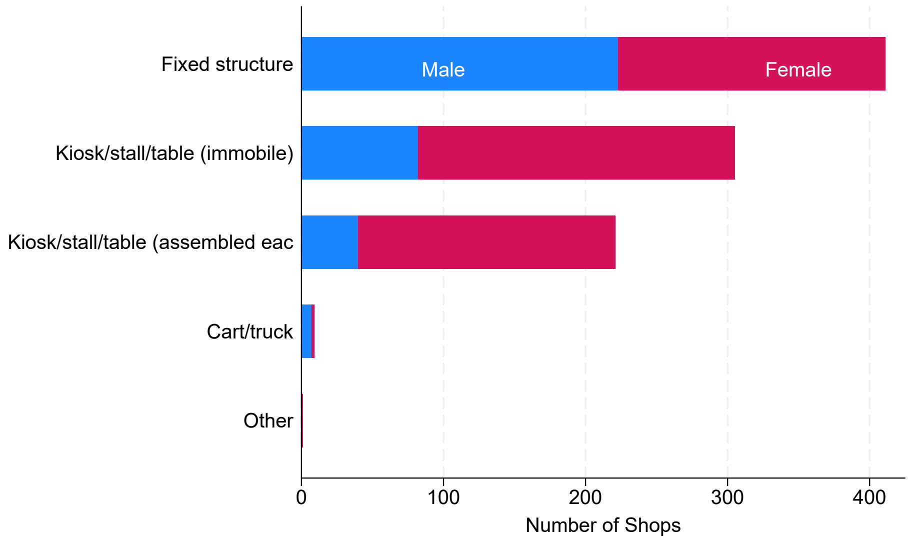
We then ran an eight-month panel survey to follow the shops surveyed during this in-person baseline. After an initial round where we either surveyed shops in-person or asked them to fill out self-surveys, we moved to a phone format for the remaining 7 months. Each shop was assigned to one of four "clusters" 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

Figure 3: Geo-coordinates of Shops in the Sample



Note: Each dot is a shop included in the in the high frequency panel.

Figure 4: Sample Counts by Shop Structure and Gender

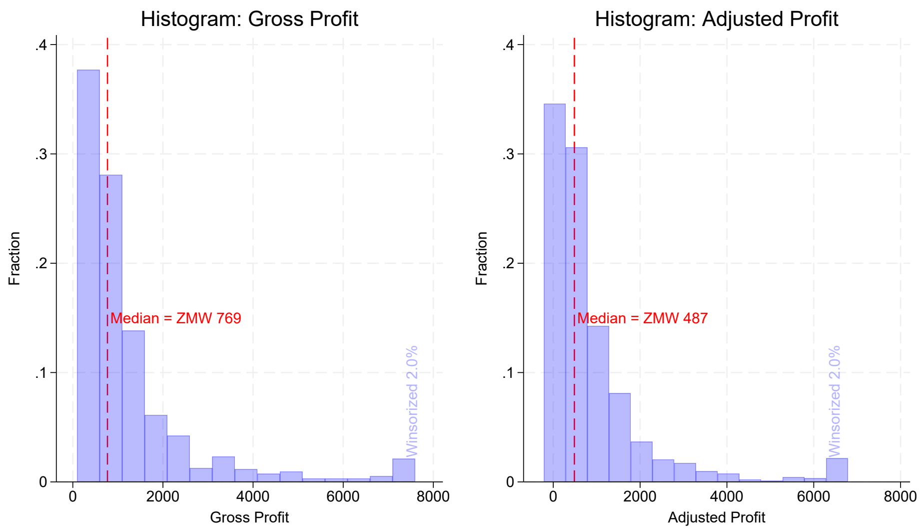


sold, number of customers, and cost of any goods purchased for resale. Some questions only needed to be asked only on specific days, like the level of inventory at the beginning and end of the week. Questions that needed to be asked only once per week—for example, the prices and costs of top products, 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.[8]

Figure 4 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.

[8]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 5: 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.

5.2 Practices and Outcomes in the Informal Retail Sector

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

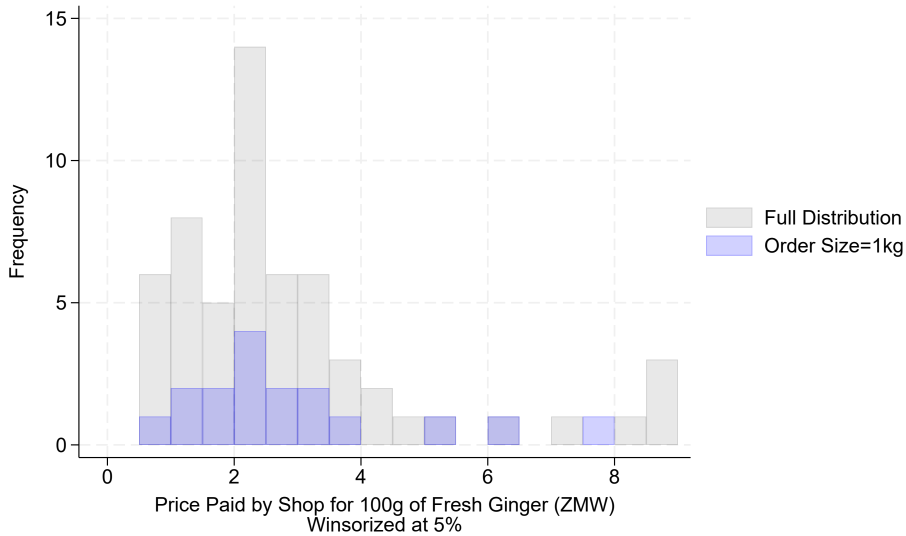
1. Gross profit: sales over one week minus the cost of the goods sold
2. Adjusted gross profit: gross profit minus the value of goods lost to theft or spoilage over the week, and minus the cost of delivery, transport, and time spent restocking during the week.[9] This is the data analog of Equation 26.

We average both across all 8 rounds and Winsorize at the 2 percent level. Figure 5 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, a significant number earn more than 4 times as much.

The model suggests many differences in practices that could explain the differences in outcomes. The first part of the model implies productive shops choose better-selling inventory, and buy it at lower prices. Differences in inventory and order prices appear in even the narrowest and most mundane niches of Lusaka's retail sector. For example, Figure 6 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.

[9]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 6: 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.

Figure 7 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.[10] 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.

Also consistent with the model, 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.

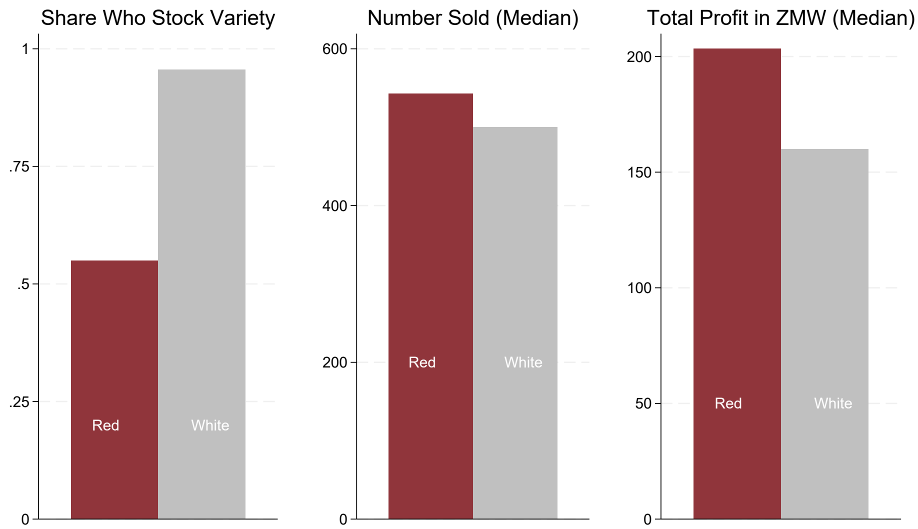
The second part of the model implies that productive shops have mastered the logistics of sourcing large quantities of goods from distant suppliers. Figure 8 shows that there is also wide variation in the logistical complexity of different shops' operations. We restrict attention to shops that report soft drinks as their most important product to ensure we are comparing stores in a roughly similar niche. The left-hand panel plots a histogram of the number of sources for inventory (the number of regular suppliers and markets).[11] While most shops buy from only two or three sources, there are some with four, six, or even 10 sources.

More sources generally implies a greater variety of products. But independent of the number of sources, there is also variation in the distance to the source of the main product (soft

[10] During the survey respondents were given the option of choosing between white, yellow, and red onions. It was subsequently discovered that genuine white onions 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.

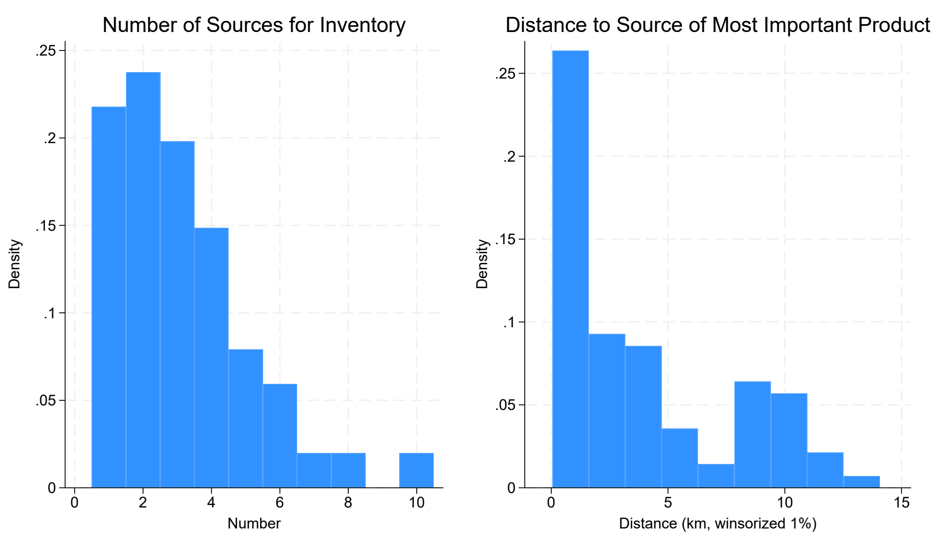
[11] There are two broad categories of sources for inventory in our sample: "regular suppliers," described as "people or businesses that you have bought from many times over the past year, and expect to buy from in the near future"; or "markets," described as "visiting a market and buying from whoever has what you are looking for." Many shops buy from both regular suppliers and markets.

Figure 7: Sellers Choose Different Goods (Variety of Onions) and Reap Different Outcomes



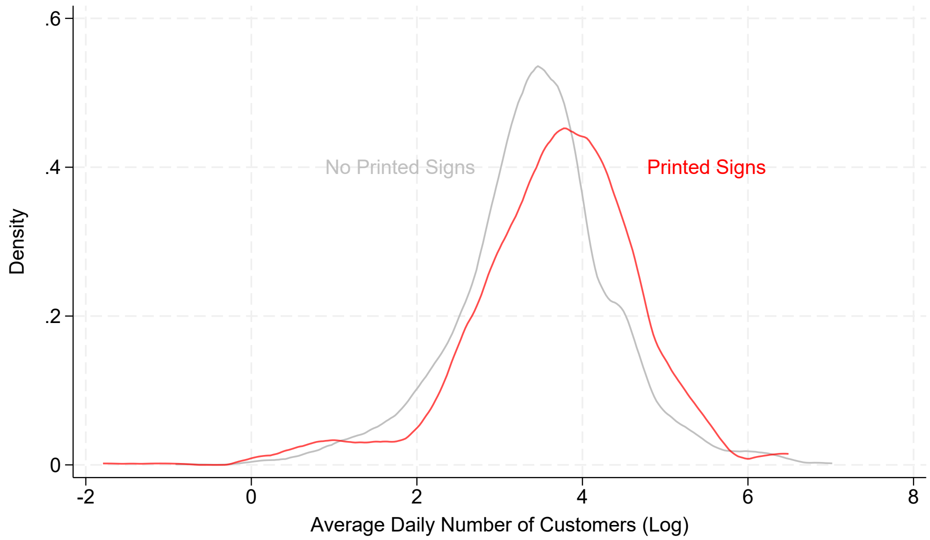
Note: Sellers who report selling onions were asked about all varieties that they stock. All data were collected from mid-January through early March during the in-person baseline survey conducted. "Number Sold" and "Total Profit" are conditional on stocking the variety, and refer to sales and profits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

Figure 8: Number of Sources and Distance to Source of Main Product (Soft Drink-Sellers)



Note: Both figures are restricted to the subsample reporting that their most important good is soft drinks. The left-hand panel shows the distribution of shops across the total number of sources for inventory (for all goods, not just the main good). A source could be either a regular supplier or a market where the shop finds whatever seller is best (cheapest, highest quality, best variety, etc.). The right-hand panel shows the distance to the source of the most important product.

Figure 9: There is Wide Variation in Daily Customers, but Shops with Printed Signs Attract More



Note:

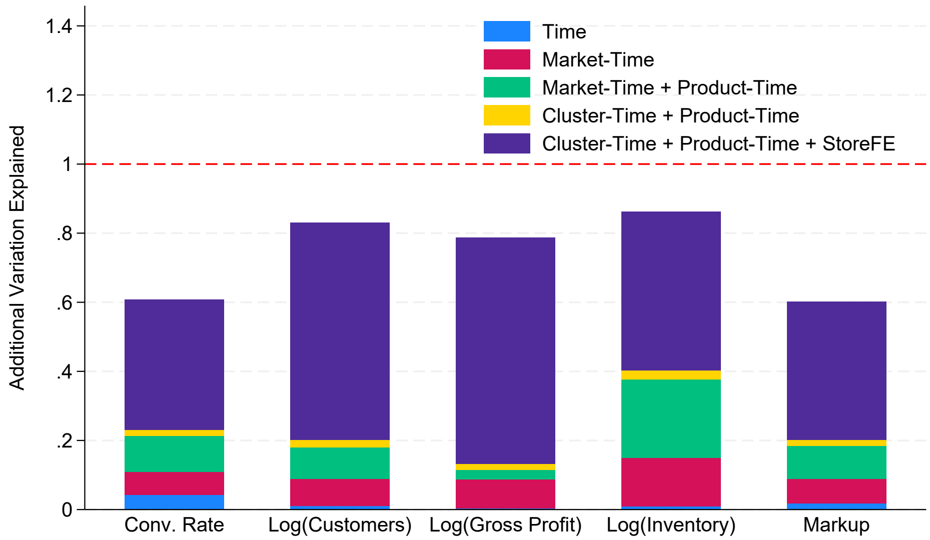
drinks). The right-hand panel shows that although the modal shop travels less than 2 kilometers, some shops travel over 5 kilometers. Shops that are able to buy from many sources at great distances are likely to have relatively sophisticated logistical operations that let them maintain a larger and richer inventory.

The third part of the model implies that productive shops are able to get more customers in the door. Figure 9 shows there is enormous variation across shops in the average number of customers per day.[12] 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).

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 10 shows the incremental variation explained in a series of key outcomes as we add additional 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. Each layer of each bar shows the additional variation explained by that model. The full model, which includes store fixed-effects, explains most of the variation for all outcomes. In all cases the final model, which includes store-fixed effects, explains more than twice as much variation as the next-best model. Though these fixed-effects are not the same as productivity—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 there is at least the potential for any or all three measures of productivity to play a major role in

[12]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 10: 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.

explaining which shops succeed or fail.

6 Estimates

6.1 Marketing Productivity

Before applying the two-step procedure to estimate marketing productivity, we estimate different versions of Equation 44 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 4.3. 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

Table 1: Marketing Productivity, Part 1: What Drives Customer to Choose One Shop Over Another? (Equations 44 45)

	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 44 excluding the fixed effect. Columns 5 and 6 show estimates of Equation 45. 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.

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 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. (Whether such an advertising strategy exists but is unknown or unused by these small retailers remains to be seen.)

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 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 (45). Although our preferred specification is Column 5, we extract the fixed effect from all three to verify that the final measure of marketing productivity is robust to different assumptions about the first stage. Columns 1—3 of Table 2 reports estimates of the second-stage estimation (Equation 46). 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 productivity, the residual of these regressions is an unbiased estimate of marketing productivity. The last three columns of the table show the correlation of this estimate across all three versions of

Table 2: Marketing Productivity, Part 2: Isolating Marketing Productivity from Average Prices and Fixed Advertising Investments (Equation 46)

	Stage 2 Estimation			Correlation Across Versions		
	(1) Version 1	(2) Version 2	(3) Version 3	(4) Version 1	(5) Version 2	(6) Version 3
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 46. 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 productivity arising from these three versions (based on simple OLS regressions).

the estimation. Versions 1 and 2, both based on Equation 45 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).

6.2 Logistical Productivity

Table 3 reports estimates of Equation 36, the estimating equation for logistical productivity. 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).[13] Column 5 estimates the enhanced model that allows depreciation to vary with the time between restocking (see Appendix B).

The capital coefficient is small and insignificant in Column 1. But it becomes large and significant (roughly 0.55) after instrumenting, consistent with a downward bias in OLS caused by

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

measurement error. The labor coefficient is consistently roughly 0.45 in Columns 2—4. Added together they imply roughly constant returns, implying that a doubling of capital and labor will allow the shop to source and stock twice as many goods. 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 0.92, though we cannot reject constant returns to scale.

Table 3: Production Elasticities (Logistical Productivity 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 36. 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 36 implies is an unbiased estimate of logistical productivity.

6.3 Inventory Choice Productivity

Inventory choice productivity is estimated as a fixed effect that is isolated using a nonparametric control function. Since the coefficients from the estimation do not necessarily represent structural parameters, it is not obvious how to interpret them. Instead we estimate 5 different versions of Equation 38 to confirm that the resulting versions of inventory choice productivity are similar. The versions are as follow:

1. Baseline specification—control function is linear in the variables listed in Equation 38
2. Controls for fully-interacted quadratic in the variables listed in Equation 38

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 productivity.

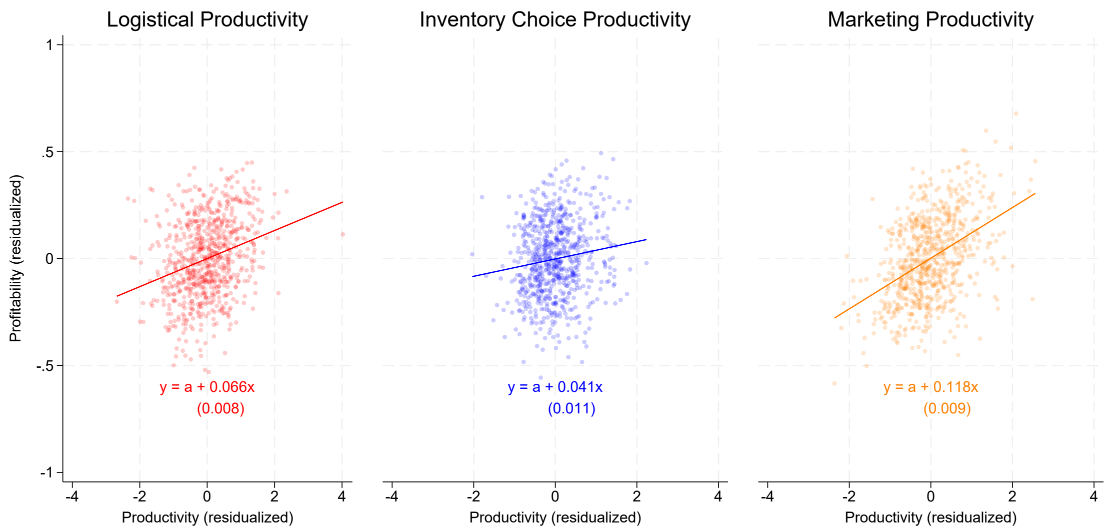
Table 4: Correlation Between Versions of Inventory Choice Productivity

	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 marketing productivity described in the text.

7 Results

Figure 11: Productivity is Correlated with Profitability



Note: Each panel shows the partial scatter plot (and line of best fit) between adjusted profit, as defined in Section 5.2, and one measure of productivity. 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 measures of productivity and a set of product-type fixed effects. The equations below the scatter plot show the slope and standard error.

7.1 Validating the Measures

Before exploring how our measures of productivity add value beyond a standard productivity estimator, we first test some basic predictions that would hold if they are valid.

The most basic is whether shops that are more productive also have higher adjusted profit, as defined in Section 5.2. Section 2.4 showed that all three measures of productivity should independently increase the level of adjusted profit. Unlike in the model, actual adjusted profit can be negative, making it difficult to define a convenient rescaling transformation like the natural logarithm.[14] We instead use the cross-round percentile of adjusted profit, which gives the relative ranking of the profit 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 profitability against one measure of productivity after partialing out the effect of the other measure of productivity and a set of broad product-type fixed-effects.[15] All measures are strongly and independently correlated with profitability.

[14]The inverse hyperbolic sine transformation creates an artificial bimodality in the distribution

[15]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, and Miscellaneous.

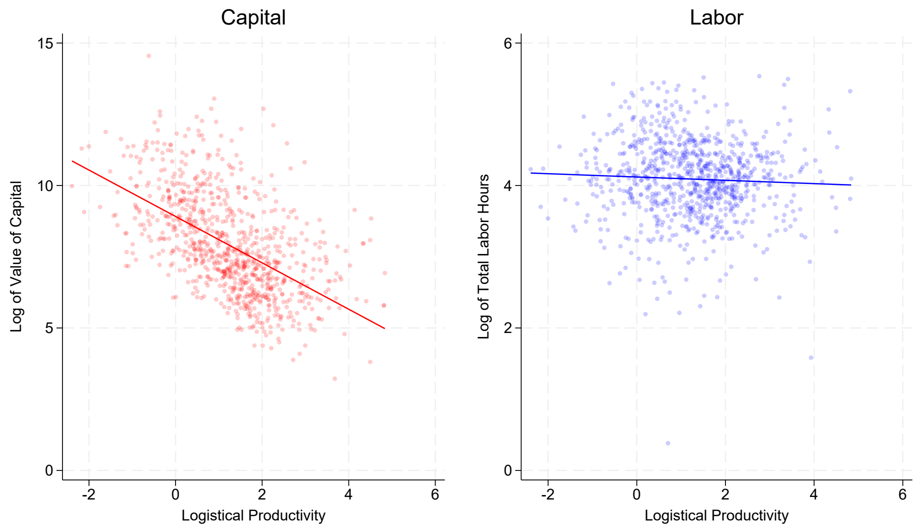
One objection to Figure 11 is that both profitability and productivity are estimated in the same data. To rule out a mechanical correlation we re-estimate the measures of productivity using only the first seven rounds of data, and measure the correlation with the cross-round percentile of adjusted profit in the eighth round. Table 5 compares the predictive power of the three measures of productivity to several binary measures of management suggested by McKenzie and Woodruff (2017) and a measure of productivity based on typical models of production. This "standard" measure of productivity is a simple fixed-effects estimator of gross output (valued at cost) that instruments for capital with investment (analogous to our measure of logistical productivity). We select this measure over more sophisticated alternatives (e.g. Ackerberg et al., 2015; Gandhi et al., 2017) because they assume productivity follows a Markov process and are thus not directly comparable to our measures (which assume all dimension of productivity is fixed).[16]

Column 1 shows that the measures of management are generally not good predictors of productivity. 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 productivity. The nine measures of management practices together explain less than 6 percent of the within-product-category variation in gross profits. That is far below the 29 percent explained by the three measures of productivity (Column 2). When combined in Column 3 the measures of management and the measures of productivity together explain 33 percent of the variation, roughly the sum of the R-squared of the separate specifications. This last fact holds because the management practices are mostly uncorrelated with the measures of productivity. The partial R-squared for the two sets of variables confirms that the productivity estimates account for the vast majority of the explained variation.

Columns 4 shows that the standard measure of productivity explains far less of the variation than the three measures of productivity (8 percent as compared to 29 percent). Clearly the three measures together contain information beyond the single measure alone. But Column 5 shows that logistical productivity, which is the most similar to the standard estimator, explains by itself slightly more variation than the standard measure of productivity (10 percent versus 8 percent). Given that both estimators have the same right-hand side, the left-hand side variable derived from the model is driving the difference in estimates (and their predictive power). While the standard estimator uses only gross output on the left-hand side, Equation 16 uses gross output plus the value of unsold inventory lost to spoilage and other depreciation. This model-derived outcome evidently leads to an estimates of productivity that is more informative about future profits. Finally, Column 6 runs a "horse race" regression with both measures. Of the two, logistical productivity retains a positive correlation while the coefficient on standard productivity flips signs. This is not surprising given that they are highly correlated with

[16]It is straightforward to generalize logistical productivity to follow an auto-regressive process, and apply an appropriate estimator, as long as productivity is fixed across restocking events within a period.

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



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

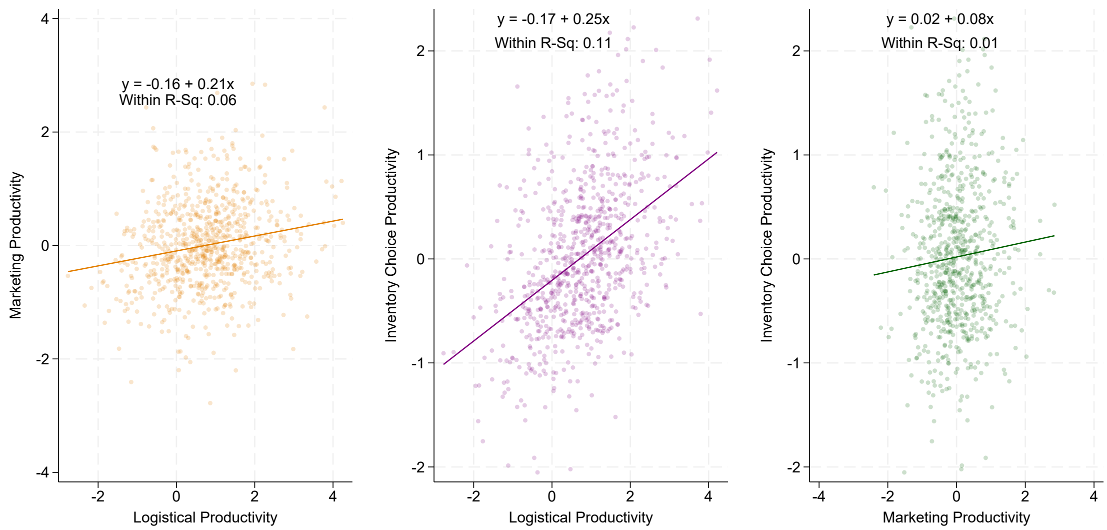
one another, but logistical productivity is slightly better correlated with the outcome.

One final validation of the measures comes from Proposition 4, which implies that a firm with higher logistical productivity may actually choose a lower level of capital (the opposite of what is predicted by the standard model of production in manufacturing). Figure 12 shows that our estimates are consistent with this prediction. Logistical productivity 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.

7.2 Standard Methods Might Yield Inaccurate Conclusions

Given that our estimates pass these basic tests, we can turn to the key question for policy-makers: how well-correlated are these measures? Section 3 showed that many "standard" predictions about optimal interventions fail when the three dimensions of productivity are imperfectly correlated. Figure 13 shows that the correlation between logistical and inventory choice productivity 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.

Figure 13: The Correlation Between the Three Dimensions of Productivity is Low

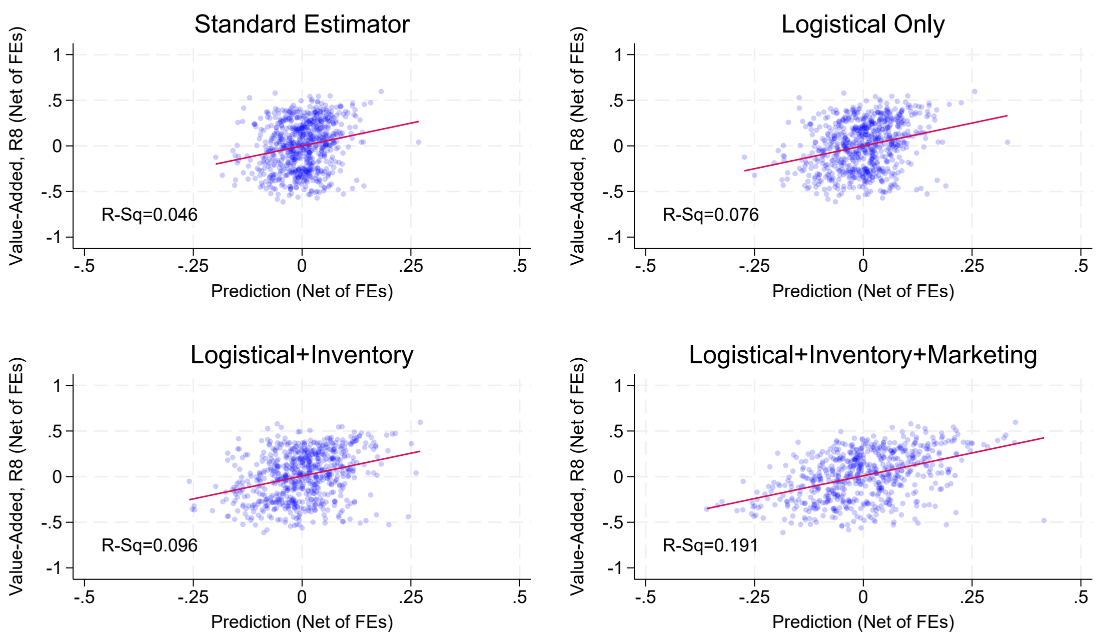


Note: Each plot presents the scatter plot, line of best fit, regression equation, and R-squared of one measure against a different measure. All pairwise regressions control for broad product category dummies.

This result implies that a policy targeting entrepreneurs deemed highly productive based on standard estimators—even if those estimators yield relatively unbiased estimates—would at best be targeting entrepreneurs with high logistical productivity. Since many would have low productivity in other dimensions, these firms would not even necessarily be the most profitable. Figure 14 illustrates this fact by showing a scatter plot of the percentile of adjusted profit in Round 8 against the fitted values from a regression of this outcome on different measures of productivity (after partialing out product-category fixed effects). The top-left panel uses predicted values from a regression of the outcome on the standard estimator. The top-right panel uses the analogous prediction based on logistical productivity. The bottom panels show specifications that incrementally add the other two dimensions of productivity. The figures show that the dispersion around the line of fit best shrinks visibly as we move to a model using all three measures of productivity. At the least, ignoring the other dimensions of productivity is throwing away useful information.

But Section 3 implies that, when considering a specific policy intervention, the consequences of ignoring the other dimensions of productivity is potentially far more serious. A policy that granted additional inventory or capital, for example, might have little impact on profits while increasing waste. Meanwhile, an econometrician studying this industry might conclude there is substantial misallocation in the factors of production because shops with "high productivity" might have less capital than shops with lower productivity. Yet when the dimensions of

Figure 14: The Three Separate Measures Have More Out-of-Sample Predictive Power than the Most Comparable Standard Estimator



Note: Each panel plots the percentile of adjusted profit in Round 8 against the fitted values from a regression of this outcome on different measures of productivity (after partialing out product category fixed effects). The text in the lower-left corner of each panel shows the R-squared from a regression of the outcome on the measures.

productivity have a low correlation, it may be perfectly efficient for shops with high logistical productivity to need less capital.

8 Conclusion and Directions for Future Research

This paper proposes a multidimensional estimator for retail productivity, and applies it to a novel dataset collected from small retailers in Lusaka, Zambia. Our estimates are more strongly correlated with future value-added than standard measures of management and a traditional fixed-effects estimate of unidimensional productivity. The three dimensions of retail productivity are only weakly correlated, implying that any analysis or policy based on standard estimates of productivity could have perverse results.

Our results imply that the standard model of production, which was intended to describe manufacturing, can yield misleading conclusions when applied to a sector whose operations are completely different. Our results also imply that there are potential insights in opening the "black box" of productivity by designing models that are more closely linked to the actual process of production. Services now account for the bulk of employment and GDP in most developed countries. Having no model of production in these industries—education, health care, and information technology to name a few—may give a faulty perspective on the modern economy. The design of such models is a natural path for future research.

References

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

ACKERBERG, D. A., K. CAVES, AND G. FRAZER (2006): "Structural Identification of Production Functions," .

——— (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83, 2411–2451.

ARMSTRONG, M. AND J. ZHOU (2011): "Paying for Prominence," *The Economic Journal*, 121, F368–F395.

BAGWELL, K. AND G. RAMEY (1994): "Coordination Economies, Advertising, and Search Behavior in Retail Markets," *The American Economic Review*, 84, 498–517.

BLAIR, B. F. AND T. R. LEWIS (1994): "Optimal Retail Contracts with Asymmetric Information and Moral Hazard," *The Rand Journal of Economics*, 25, 284–296.

BLOOM, N., C. GENAKOS, R. SADUN, AND J. VAN REENEN (2012): “Management Practices Across Firms and Countries,” *Academy of Management Perspectives*, 26, 12–33.

BLUNDELL, R. AND S. BOND (2000): “GMM Estimation With Persistent Panel Data: An Application to Production Functions,” *Econometric Reviews*, 19, 321–340.

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

DALTON, P. S., J. RÜSCHENPÖHLER, B. URAS, AND B. ZIA (2021): “Curating Local Knowledge: Experimental Evidence from Small Retailers in Indonesia,” *Journal of the European Economic Association*.

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.

ELLICKSON, P. B., P. L. E. GRIECO, AND O. KHVASTUNOV (2020): “Measuring Competition in Spatial Retail,” *The Rand Journal of Economics*, 51, 189–232.

FAIG, M. AND B. JEREZ (2005): “A Theory of Commerce,” *Journal of Economic Theory*, 122, 60–99.

FOSTER, L., J. HALTIWANGER, AND C. J. KRIZAN (2006): “Market selection, reallocation, and restructuring in the US retail trade sector in the 1990s,” *The Review of Economics and Statistics*.

GANDHI, A., S. NAVARRO, AND D. RIVERS (2017): “On the Identification of Gross Output Production Functions,” .

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,” .

JOHNSON, J. P. (2017): “Unplanned Purchases and Retail Competition,” *The American Economic Review*, 107, 931–965.

KREMER, M. (1993): “The O-Ring Theory of Economic Development,” *The Quarterly Journal of Economics*, 108, 551–575.

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

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

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

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

SMITH, H. (2004): "Supermarket Choice and Supermarket Competition in Market Equilibrium," *The Review of Economic Studies*, 71, 235–263.

SOLOW, R. (1957): "Technical Change and the Aggregate Production Function," *The Review of Economics and Statistics*, 39, 312–320.

THOMASSEN, Ø., H. SMITH, S. SEILER, AND P. SCHIRALDI (2017): "Multi-category Competition and Market Power: A Model of Supermarket Pricing," *The American Economic Review*, 107, 2308–2351.

VARIAN, H. R. (1980): "A Model of Sales," *The American Economic Review*, 70, 651–659.

WOOLDRIDGE, J. M. (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, 104, 112–114.

YOU, W. (2021): "The Economics of Speed: The Electrification of the Streetcar System and the Decline of Mom-and-Pop Stores in Boston, 1885–1905," *American economic journal. Applied economics*, 13, 285–324.

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 Productivity		0.048***	0.053***		0.078***	0.701***
		(0.011)	(0.011)		(0.010)	(0.059)
Inventory Choice Productivity		0.050***	0.045***			
		(0.015)	(0.015)			
Marketing Productivity		0.111***	0.106***			
		(0.013)	(0.013)			
Standard Measure of Productivity				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 productivity, several binary measures of management suggested by McKenzie and Woodruff (2017), and a standard fixed-effects estimate of productivity 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 productivity (that is, the share of residual variation in a specification excluding that variable that would be explained if the variable were included). The "standard" measure of productivity is a simple fixed-effects estimator of gross output (valued at cost) that instruments for capital with investment (analogous to our measure of logistical productivity).

A Proofs of Propositions in the Main Text

A.1 Derivation of (Expected) Total Sales

Substitute Equation 8 into the expression for expected sales:

$$\mathbb{E}[Y] = \sum_{j=1}^{J} P_j \left\{ \left[1 - \Phi\left(\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J} \right) \right] x_j + \Phi\left(\frac{x_j - \tilde{\alpha}_j DS/J}{\sigma DS/J} \right) \mathbb{E}[\alpha_j DS/J \mid \alpha_j DS/J < x_j] \right\} \tag{48}$$

$$= \mu \left(\sum_{j=1}^{J} \kappa_j x_j \right) [1 - \Phi(\upsilon)] + \mu \Phi(\upsilon) \left(\sum_{j=1}^{J} \kappa_j \mathbb{E}[\alpha_j DS/J \mid \alpha_j DS/J < x_j] \right) \tag{49}$$

$$= \mu \left(\sum_{j=1}^{J} \kappa_j x_j \right) [1 - \Phi(\upsilon)] + \mu \left(\frac{1}{J} \sum_{j=1}^{J} \kappa_j \tilde{\alpha}_j DS \right) \Phi(\upsilon) - \mu\sigma DS \left(\frac{1}{J} \sum_{j=1}^{J} \kappa_j \right) \phi(\upsilon) \tag{50}$$

$$= \mu X [1 - \Phi(\upsilon)] + \mu\zeta DS \Phi(\upsilon) - \mu\sigma DS \bar{\kappa} \phi(\upsilon) \tag{51}$$

where (49) follows from (8) and the fixed markup assumption, (50) follows from the truncated expectation of a normal random variable, and (51) from applying (3) and (6). Collect terms and apply the definition of υ to form a final expression for sales within a restocking period:

$$\mathbb{E}[Y] = \mu \Big\{ X - \sigma\bar{\kappa} DS [\Phi(\upsilon)\upsilon + \phi(\upsilon)] \Big\} \tag{52}$$

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

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

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

$$0 = \sum_{j=1}^{J} [P_j \min\{x_j, DS\alpha_j/J\} - \mathbb{E}(P_j \min\{x_j, DS\alpha_j/J\})] \tag{53}$$

$$= \sum_{j=1}^{J} [P_j \min\{x_j, DS\alpha_j/J\}] - \mathbb{E}\left(\sum_{j=1}^{J} P_j \min\{x_j, DS\alpha_j/J\} \right) \tag{54}$$

$$= Y - \mathbb{E}[Y] \tag{55}$$

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

$$\mathcal{V}(U_t; \{\alpha_{jt}\}, \{P_{jt}\}) \tag{56}$$

$$= \max_{\{x_{jt}\}} \mathbb{E}\Big[\sum_{j=1}^{J} P_{jt} \min\{x_{jt}, D_t S_t \alpha_{jt}/J\} + \vartheta_t \left(\mathcal{D}(S_{t-1})U_t + B_t - \sum_{j=1}^{J} \kappa_{jt} x_{jt}\right) \tag{57}$$

$$+ \mathcal{V}(U_{t+S_t}; \{\alpha_{j,t+S_t}\}, \{P_{j,t+S_t}\})\Big] \tag{58}$$

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

$$\kappa_{jt}\vartheta_t = P_{jt}\left[1 - \Phi\left(\frac{x_{jt} - D_t S_t \tilde{\alpha}_{jt}/J}{\sigma D_t S_t/J}\right)\right] + \mathbb{E}\left[\mathcal{V}_1(U_{t+S_t}; \{\alpha_{j,t+S_t}\}, \{P_{j,t+S_t}\})\frac{\partial U_{t+S_t}}{\partial x_{jt}}\right] \tag{59}$$

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

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

Combine these conditions and apply the fixed markup assumption:

$$\vartheta_t = \mu_t \left[1 - \Phi\left(\frac{x_{jt} - D_t S_t \tilde{\alpha}_{jt}/J}{\sigma D_t S_t/J}\right)\right] + \mathcal{D}(S_t)\Phi\left(\frac{x_{jt} - \tilde{\alpha}_{jt} D_t S_t/J}{\sigma D_t S_t/J}\right)\mathbb{E}[\vartheta_{t+S_t} \mid x_{jt} \geq D_t S_t \alpha_{jt}/J] \tag{60}$$

This condition will be satisfied for any choice of x_{jt} that makes the right-hand side of Equation 60 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

$$\vartheta_t = \mu_t \left[1 - \Phi\left(\upsilon_t\right)\right] + \mathcal{D}(S_t)\Phi\left(\upsilon_t\right)\mathbb{E}\left[\vartheta_{t+S_t} \mid \upsilon_t \geq \frac{\alpha_{jt} - \tilde{\alpha}_{jt}}{\sigma}\right] \tag{61}$$

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

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

where ζ_t and $\bar{\kappa}_t$ 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 58 is equivalent to

$$\mathcal{D}(S_{t-1})U_t + B_t - \sum_{j=1}^{J} \kappa_{jt} x_{jt} = \mathcal{D}(S_{t-1})(X_t - Y_t/\mu_t) + B_t - X_t \tag{62}$$

$$\Rightarrow X_t \leq Z_t K_t^{\beta} L_t^{\nu} + \mathcal{D}(S_{t-1})(X_t - Y_t/\mu_t) \tag{63}$$

which is simply the constraint (15) in Section 2.2 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

$$G = X - (1-\delta)(X - Y/\mu) \tag{64}$$

$$= X - (1-\delta)\frac{\sigma D \bar{\kappa}}{Q}\left[\Phi\left(\upsilon\right)\upsilon + \phi(\upsilon)\right] \tag{65}$$

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

Then the Lagrangian is

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

The first-order conditions are

$$\vartheta\frac{\partial G}{\partial X} = \left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]Q \tag{66}$$

$$\vartheta\frac{\partial G}{\partial Q} = \left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f \tag{67}$$

Now suppose that the constraint does not bind and $\vartheta = 0$. Then Equation 66 becomes

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

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 67 implies that $f = 0$, a contradiction. Then it must be that the constraint always binds.

A.4 Proof of Proposition 3

Divide (67) by (66) and rearrange:

$$\frac{\frac{\partial G}{\partial Q}}{\frac{\partial G}{\partial X}} = \frac{\left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f}{\left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]Q} \tag{68}$$

$$\frac{(1-\delta)\left\{Y/\mu - X + \Phi\left(\upsilon\right)X\right\}}{(1-\delta)\Phi\left(\upsilon\right) - 1} = \frac{\left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]X - f}{\left[\mu\left\{1 - \Phi\left(\upsilon\right)\right\} + (1-\delta)\Phi\left(\upsilon\right) - 1\right]} \tag{69}$$

$$\frac{ZK^{\beta}L^{\nu} + \left[(1-\delta)\Phi\left(\upsilon\right) - 1\right]X}{(1-\delta)\Phi\left(\upsilon\right) - 1} = X - \frac{f}{\mu - 1 - \left\{\mu - (1-\delta)\right\}\Phi\left(\upsilon\right)} \tag{70}$$

$$X - \frac{ZK^{\beta}L^{\nu}}{1 - (1-\delta)\Phi\left(\upsilon\right)} = X - \frac{f}{\mu - 1 - \left\{\mu - (1-\delta)\right\}\Phi\left(\upsilon\right)} \tag{71}$$

$$\frac{f}{\mu\left[1 - \Phi\left(\upsilon\right)\right] - \left[1 - (1-\delta)\Phi\left(\upsilon\right)\right]} = \frac{ZK^{\beta}L^{\nu}}{1 - (1-\delta)\Phi\left(\upsilon\right)} \tag{72}$$

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

$$\frac{m - (m+\delta)\Phi(\upsilon)}{f} = \frac{1 - (1-\delta)\Phi(\upsilon)}{B} \tag{73}$$

$$mB - (m+\delta)B\Phi(\upsilon) = f - (1-\delta)f\Phi(\upsilon) \tag{74}$$

$$mB - f = [(m+\delta)B - (1-\delta)f]\Phi(\upsilon) \tag{75}$$

$$\Phi^{-1}\left(\frac{mB - f}{(m+\delta)B - (1-\delta)f}\right) = \upsilon \tag{76}$$

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

$$\max_{Q^*, X^*} Q^* \left\{Y(X^*, Q^*) - \left[X^* - (1-\delta)(X^* - Y(X^*, Q^*)/\mu)\right] - f\right\} \tag{77}$$

$$+ \vartheta[(1-\delta)(X^* - Y(X^*, Q^*)/\mu) + ZK^{\beta}L^{\nu} - X^*] - p^K K - p^L L \tag{78}$$

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

$$\vartheta\beta\frac{ZK^{\beta}L^{\nu}}{K} = p^K \tag{79}$$

Subbing (66) into (67) shows that

$$-\vartheta\frac{\partial G}{\partial Q} = [\mu\{1 - \Phi(\upsilon)\} + (1-\delta)\Phi(\upsilon) - 1]\, Q\frac{X}{Q} - f \tag{80}$$

$$= \left(-\vartheta\frac{\partial G}{\partial X}\right)\frac{X}{Q} - f \tag{81}$$

$$\Rightarrow \vartheta = \frac{f}{\frac{\partial G}{\partial Q} - \frac{\partial G}{\partial X}\frac{X}{Q}} \tag{82}$$

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

$$\frac{\partial G}{\partial Q} = (1-\delta)\frac{1}{Q}\{Y/\mu - X + \Phi(\upsilon)X\} \tag{83}$$

$$= \frac{1}{Q}\{(1-\delta)(Y/\mu - X) + X - X + (1-\delta)\Phi(\upsilon)X\} \tag{84}$$

$$= \frac{1}{Q}\left\{ZK^{\beta}L^{\nu} + ((1-\delta)\Phi(\upsilon) - 1)X\right\} \tag{85}$$

$$= \frac{1}{Q}\left\{ZK^{\beta}L^{\nu} + \frac{\partial G}{\partial X}X\right\} \tag{86}$$

Subbing this back into (82) yields

$$\vartheta = \frac{fQ}{ZK^{\beta}L^{\nu}} \tag{87}$$

which substituted into (79) yields

$$K = \beta\frac{fQ}{p^K} \tag{88}$$

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

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

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

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

By definition

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

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

$$x_j = \frac{\tilde{\alpha}_j DS}{J} P_j^{-\mathcal{K}} + \frac{\sigma DS}{J} P_j^{-\mathcal{K}} \Phi^{-1}\left(1 - \frac{\lambda}{\mu}\right) \tag{89}$$

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

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

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

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

Redefine

$$\upsilon = \frac{X - \zeta DS}{\sigma DS \bar{\kappa}} \mu^{-\mathcal{K}} \tag{90}$$

Then algebra parallel to that in the main text yields

$$Y \approx \mu \Big\{ X - \sigma \bar{\kappa} DS \mu^{-\mathcal{K}} \left[\Phi\left(\upsilon\right) \upsilon + \phi\left(\upsilon\right)\right] \Big\} \tag{91}$$

as J becomes large. Subbing Equation 91 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

$$\upsilon = \Phi^{-1}\left(\frac{mB - f}{(m+\delta)B - (1-\delta)f}\right) \tag{92}$$

Subbing Equation 92 into Equation 91 and solving for W_{nt} as defined in Section 4 implies a nonparametric estimating equation identical to Equation 38. The only difference is that in the final measure of inventory choice productivity

$$\hat{\iota}_n \approx \log \sigma + \log \bar{\kappa}$$

the average cost $\bar{\kappa}$ 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

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

Subject to:

$$X \leq \mathcal{D}(Q)(X - Y/\mu) + HZK^{\beta}L^{\nu} \tag{93}$$

which replaces the constraint (15) with (93). 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

$$\vartheta \frac{\partial G}{\partial X} = \left[\mu \left\{1 - \Phi\left(\upsilon\right)\right\} + \mathcal{D}(Q)\Phi\left(\upsilon\right) - 1\right] Q \tag{94}$$

$$\vartheta \frac{\partial G}{\partial Q} = \left[\mu \left\{1 - \Phi\left(\upsilon\right)\right\} + \mathcal{D}(Q)\Phi\left(\upsilon\right) - 1\right] X - f + \mathcal{D}'(Q)\sigma\bar{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big] \tag{95}$$

The proof of Proposition 2 from Appendix A.3 no longer holds because the term in (95) 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 (95) by (94). This ratio yields the equation

$$\frac{f - \mathcal{D}'(Q)\sigma\bar{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big]}{\mu\left[1-\Phi\left(\upsilon\right)\right] - \left[1-\mathcal{D}(Q)\Phi\left(\upsilon\right)\right]} = \frac{B + \mathcal{D}'(Q)\sigma\bar{\kappa}D\Big[\Phi(\upsilon)\upsilon + \phi(\upsilon)\Big]}{1-\mathcal{D}(Q)\Phi\left(\upsilon\right)} \tag{96}$$

where $B = HZK^{\beta}L^{\nu}$ This expression reduces to (72) 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 (96) reduces to

$$\frac{\mu\Big[1-\Phi(\upsilon)\Big]}{1-\mathcal{D}(Q)\Phi(\upsilon)} = \frac{f+B}{B+\mathcal{E}(Q)\Big[X-B\Big]} \tag{97}$$

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 2.

B.3 Dynamic Relationship Capital

$$d_{nt} = \omega r_{nt} - \omega\left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{98}$$

$$= \omega\Big(\bar{r}_n + \delta a_{nt} + \rho r_{n,t-1} + e_{nt}\Big) - \omega\left[G\tilde{p}_{nt} + (1-G)\hat{p}_n\right] + \phi_{gt} \tag{99}$$

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

$$d_{nt} = \omega r_{nt} - \omega\left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{100}$$

$$= \omega\Big(\bar{r}_n + \delta a_{nt} + \rho r_{n,t-1} + e_{nt}\Big) - \omega\left[G\bar{p}_{nt} + (1-G)\hat{p}_n + G\nu_{nt}\right] + \phi_{gt} \tag{101}$$

Combine (101) with a one-period lag of (100) :

$$d_{nt} = \omega\Big(\bar{r}_n + \delta a_{nt} + e_{nt}\Big) + \rho\Big(d_{n,t-1} + \omega\left[G\bar{p}_{n,t-1} + (1-G)\hat{p}_n + G\nu_{n,t-1}\right] - \phi_{g,t-1}\Big) \tag{102}$$

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

$$= \omega\left[\bar{r}_n - (1-G)(1-\rho)\hat{p}_n\right] + (\phi_{gt} - \rho\phi_{g,t-1}) + \rho d_{n,t-1} \tag{104}$$

$$- \omega G\Big(\bar{p}_{nt} - \rho\bar{p}_{n,t-1}\Big) + \omega\delta a_{nt} + \omega\Big[e_{nt} - G\big(\nu_{nt} - \rho\nu_{n,t-1}\big)\Big] \tag{105}$$

$$= \pi_{0,n} + \bar{\phi}_{gt} + \pi_1 d_{n,t-1} + \pi_2\bar{p}_{nt} + \pi_3\bar{p}_{n,t-1} + \pi_4 a_{nt} + \varepsilon_{nt} \tag{106}$$

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 (106) yields

$$\ddot{d}_{nt} = \ddot{\bar{\phi}}_{gt} + \pi_1\ddot{d}_{n,t-1} + \pi_2\ddot{\bar{p}}_{nt} + \pi_3\ddot{\bar{p}}_{n,t-1} + \ddot{\pi}_4 a_{nt} + \ddot{\varepsilon}_{nt} \tag{107}$$

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

$$\hat{\pi}_{0,n} = \frac{1}{T-2}\sum_{t=2}^{T-1}\left[\breve{d}_{nt} - \hat{\pi}_1\breve{d}_{n,t-1} - \hat{\pi}_2\breve{\bar{p}}_{nt} - \hat{\pi}_3\breve{\bar{p}}_{n,t-1} - \hat{\pi}_4\breve{a}_{nt}\right] \tag{108}$$

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.

C Data Appendix

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

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.

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]

[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).

[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.

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 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 46.

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.

IGC

theigc.org