PEDL Research Papers

This research was partly or entirely supported by funding from the research initiative Private Enterprise Development in Low-Income Countries (PEDL), a Foreign, Commonwealth & Development Office (FCDO) funded programme run by the Centre for Economic Policy Research (CEPR).

NBER WORKING PAPER SERIES

MEASURING PRODUCTIVITY:
LESSONS FROM TAILORED SURVEYS AND PRODUCTIVITY BENCHMARKING

David Atkin
Amit Khandelwal
Adam Osman

Working Paper 25471
http://www.nber.org/papers/w25471

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2019

We thank AbdelRahman Nagy and the Egypt field team. We acknowledge generous funding from the International Growth Centre, Private Enterprise Development for Low-Income Countries, Innovations for Poverty Action, Economic Growth Center at Yale University, McMillan Center at Yale University and the Jerome A. Chazen Institute for Global Business at Columbia Business School. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking
David Atkin, Amit Khandelwal, and Adam Osman
NBER Working Paper No. 25471
January 2019
JEL No. D2,F0

ABSTRACT

We use tailored surveys and benchmarking in the flat-weave rug industry to better understand the shortcomings of standard productivity measures. TFPQ performs poorly because of variation in product specifications across firms. Controlling for specifications aligns TFPQ with lab benchmarks. We also collect quality metrics to construct quality productivity (the ability to produce quality given inputs) and find substantial dispersion across firms. This motivates interest in multi-dimensional productivity, or capability. As quality productivity is negatively correlated with TFPQ, TFPR may perform better at capturing capabilities in settings where better firms make products with more demanding specifications that have greater input requirements.

David Atkin
Department of Economics, E52-550
MIT
50 Memorial Drive
Cambridge, MA 02142
and NBER
atkin@mit.edu

Amit Khandelwal
Graduate School of Business
Columbia University
Uris Hall 606, 3022 Broadway
New York, NY 10027
and NBER
ak2796@columbia.edu

Adam Osman
Department of Economics
University of Illinois at Urbana-Champaign
109 David Kinley Hall, 1407 W. Gregory
Urbana, IL 61801
aosman@illinois.edu

1 Introduction

Economists have long recognized that the reason some countries are richer than others is not primarily due to differences in their endowments or resources, but because of how effectively their firms deploy those resources to generate economic activity. This prompts a set of obvious questions. How large are productivity differences across firms? What drives this dispersion? What policies are most effective at raising productivity? Answering these questions is an active area of research (see the review by Syverson, 2011) and central to this goal is the ability to accurately measure the productivity of firms.

What the researcher typically wants is a measure of physical output conditional on physical inputs, termed quantity-based productivity (TFPQ). This requires data on input and output quantities that are not typically available. In cases where these data are available, quantities are likely measured with substantial error since they cannot be easily read off accounting statements. Even if well measured, product specifications and quality levels can vary dramatically across firms and within firms across product lines—variation that is not well captured by disaggregated product categories in typical administrative datasets. This makes it difficult to both measure productivity for firms that produce many varieties or to compare productivity across firms making different varieties. Multi-product firms pose further challenges since output and input mixes vary even more widely across products than within.

As most firm-level datasets only provide expenditure and revenue data, much of the literature relies on revenue-based productivity (TFPR) measures that also capture differences in markups and quality across firms. However, if a firm's capabilities come from its ability to produce both quality and quantity, TFPR may be closer to the object of interest even though it confounds forces unrelated to productivity.

The existing literature has pursued various approaches to understand and to mitigate these measurement issues by drawing upon additional data (e.g., prices) alongside (often strong) identifying assumptions.[1]

We take a different approach. We develop tailored surveys focusing on a specific industry—flat-weave rugs—that directly address many of these measurement issues through the combination of detailed product specifications and external assessments of quality. The surveys allow us to calculate not only quantity productivity (the ability to produce quantity with a given set of inputs), but also quality productivity (the ability to produce quality with a given set of inputs) and capabilities (the combination of the two, essentially a TFPQ measure using quality-adjusted

[1] See de Loecker and Goldberg (2014) for a review of identification assumptions and measurement issues in production function estimation.

quantities).[2]

To better understand the shortcomings of standard productivity measures and potential remedies, we compare survey-based productivity measures to productivity benchmarking exercises which we argue are closest to true productivity. We find that standard TFPQ performs poorly at measuring quantity productivity, shows excessive dispersion across firms, and is inversely correlated with quality productivity. Controlling for product specifications—effectively making apples-to-apples comparisons—goes a long way towards remedying those deficiencies. Although TFPR does better than TFPQ at capturing broad capabilities, it performs worse than methods that combine survey information with explicit quality measures.

2 Survey Design and Data

Our data come from surveys we designed and administered on 219 rug-making firms in Fowa, Egypt. These firms produce a type of kilim rug called "duble" using double-treddle foot powered looms. As part of a randomized experiment exploring the impact of exporting we recruited all firms with 1-5 workers making this type of rug.

Rug producers receive orders with a particular set of specifications that include the design, thread types, and thread count. Producers prepare the appropriate inputs, install the threads on the loom and weave the rug. Although duble rugs are already a subset of a 10-digit HS-product code, there are many varieties (we observe 435 unique combinations of specifications).

In addition to rugs having different specifications, rugs also differ in quality. Unlike specifications—codifiable attributes of the rug that are typically chosen by the buyer—quality depends on weaving technique and is difficult to codify or contract on. For example, how flat the rug lies is determined by how skillfully the firm installs the thread on the loom, and whether the threads are held correctly while weaving.

We created a survey instrument to address the measurement issues noted above in contexts where output varies in quality and firms produce many varieties. We administered 6 rounds of surveys at the product-line level capturing the rug produced in the prior month. (As production runs last longer than a month in this industry, this was almost always a single variety of rug.) These surveys recorded detailed rug specifications; prices and quantities of all inputs and outputs; and labor hours spent on production and preparation activities. We also hired an independent quality assessor who graded each rug that the firm was working on at the time of the survey across 11 different dimensions (grading on a 1 to 5 scale).[3]

[2] Hallak and Sivadasan (2013) also explore multidimensional firm productivity.

[3] The dimensions are: corners, waviness, packedness, weight, touch, warp thread tightness, firmness, design accuracy, warp thread packedness, inputs, and loom.

Additionally, we set up a controlled laboratory in a rented space where all firms were paid a flat fee for their head weaver to produce a $0.98m^2$ rug with identical specifications using identical material inputs and capital equipment we provided. We recorded dimensions, weight and time taken to weave the resulting rug, and sent the rugs to be scored anonymously by both our quality assessor and a local professor of handicraft science. Atkin, Khandelwal and Osman (2017) provide further details on the sample, rug production, and the laboratory.

Online Appendix Table 1 provides summary statistics of the survey and lab, and Online Appendix Table 2 shows that our six product specifications (thread type, thread count, design difficulty, number of colors, market segment, duble subcategory) capture rug varieties relatively well—specifications explain about half the variation in prices, output and revenue, and dimensions such as thread count and type have the expected signs.

3 Measuring Productivity

We calculate several productivity measures from the survey data. The first measure we call "unadjusted" productivity because, as in the existing literature, it does not adjust for the fact that different firms produce rugs with different specifications (i.e., different varieties). We estimate unadjusted TFPQ (ϕ_u) from a Cobb-Douglas production function:

$$x = \phi_u l^{\alpha_l} k^{\alpha_k} e^{\epsilon} \tag{1}$$

where x is output in square meters, l is labor hours, k is capital (number of looms), and ϵ is measurement error.[4] For transparency, we estimate (1) in logs over every firm-round observation using OLS and recover ϕ_u by exponentiating the residual. The Online Appendix replicates the analysis estimating (1) using a control function.

Although this formulation is standard, a number of features reduce measurement concerns compared to other settings. First, we observe quantities of x, l and k rather than revenues and expenditures. Second, given the simple technology there are essentially no other inputs used in production (e.g., no accounting, logistics, human resources). Third, we recorded the inputs used for each specific rug produced so there is no error in allocating inputs to outputs.

Our second measure, "specification-adjusted" productivity, is more novel because it controls for differences in the variety mix across firms that may make standard TFPQ measures misleading. To guide our specification adjustment, we place more structure on equation (1) by assuming $\phi_u = \phi_a e^{\lambda\gamma}$, where λ denotes the vector of specifications which affect how quickly a rug can be

[4] At this level of disaggregation, the production function is best characterized as Leontief in materials. The unit of analysis is the firm-round level.

produced (e.g., a high thread count rug requires more labor and capital inputs) and γ are parameters to estimate. ϕ_a is specification-adjusted TFPQ that is recovered from estimating the production function conditioning on the six specification controls.

These two measures essentially capture how many labor hours firms require to produce rug quantity, potentially controlling for the specifications of the rug; we call these *quantity* productivity. While the literature typically explores a single dimension of productivity, as discussed above, similar specification rugs also vary substantially in quality. Thus, there is a second dimension of productivity that also raises revenues:[5] the skill of a firm at producing quality from a given set of inputs. We term this *quality productivity*, or TFPZ.

It is necessary to construct a quality index in such a way that quality and quantity productivity estimates can be compared and aggregated. To do so, we let the consumers' relative valuation of quantity and quality guide us. For simplicity, we make the assumption that consumers have CES demands between rugs and another good y, where consumers trade off the quality and quantity of rugs as follows: $U = ((\Pi_j q_j^{\theta_j} x)^{\frac{\sigma-1}{\sigma}} + (y)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}}$. The vector θ determines the trade off between quantity and the eleven dimensions of quality $\mathbf{q}$ indexed by j. This implies demands:

$$\ln x = (\sigma - 1) \sum_j \theta_j \ln q_j - \sigma \ln p + c \tag{2}$$

where p is the price of the variety and c is a function of total expenditure and the rug price index that is common across varieties. We recover the θ_js by regressing $(\ln x + \sigma \ln p)/(\sigma - 1)$ on our eleven quality metrics and setting $\sigma = 2.74$ based on Broda and Weinstein (2006).[6]

We next conjecture a production function for producing consumers' valued quality, $\Pi_j q_j^{\theta_j}$, with the same functional form as the quantity production function:

$$\Pi_j q_j^{\theta_j} = \zeta_u l^{\beta_l} k^{\beta_k} e^{\varepsilon} \tag{3}$$

where ζ_u is the residual after conditioning on labor and capital inputs.[7] We assume $\zeta_u = \zeta_a e^{\lambda\delta}$, with the residual increasing in the firm's quality productivity, ζ_a, and allowing the residual to also depend on specifications (for example, ensuring high quality for a high thread count rug may require more inputs than for a low thread count rug).[8] Thus, as with quantity productivity, we estimate two variants of quality productivity: unadjusted TFPZ (ζ_u) and specification-adjusted

[5] Atkin, Khandelwal and Osman (2017) shows that prices conditional on specifications increase in quality.

[6] This is their average elasticity estimate within the six-digit HS category for these rugs (HS 570231).

[7] In principle we could also adjust inputs with measures of input quality, particularly worker skill.

[8] Atkin, Khandelwal and Osman (2017) implicitly assume $\beta_l = \beta_k = 0$ as skill (rather than l and k) primarily determines quality in this industry. Here, we posit similar production functions for quality and quantity.

TFPZ (ζ_a).

Given our assumptions on supply and demand, it is straightforward to aggregate quantity and quality productivity by forming a production function for the $\Pi_j q_j^{\theta_j} x$ aggregator valued by consumers ($\Pi_j q_j^{\theta_j} x = \zeta_a \phi_a e^{\lambda(\gamma+\delta)} l^{\alpha_l+\beta_l} k^{\alpha_k+\beta_k} e^{\epsilon+\varepsilon}$). The implied productivity aggregator, which we term specification-adjusted TFPC or firm capability, is the product of specification-adjusted TFPZ and TFPQ ($\zeta_a \phi_a$).[9] Similarly, unadjusted TFPC is $\zeta_u \phi_u$.

Finally, we also estimate standard TFPR from equation (1) but replacing x, l and k with rug revenues, labor expenditures and the value of the capital stock, respectively.[10]

We compare the survey-based measures with productivity benchmarks from the controlled lab. The lab provides direct measures of quantity productivity in meters squared per unit input: Lab TFPQ $= .98m^2/(l_{lab}^{\hat{\alpha}_l} k_{lab}^{\hat{\alpha}_k})$, where l_{lab} is the hours taken to produce the rug in the lab, and $k_{lab} = 1$ is the number of looms. We calculate lab quality productivity, Lab TFPZ$= \Pi_j q_{lab,j}^{\hat{\theta}_j}/(l_{lab}^{\hat{\beta}_l} k_{lab}^{\hat{\beta}_k})$, by combining the anonymized quality assessments for the lab rugs (averaging over the two experts' grades) with the $\hat{\theta}_j$s from regression (2). The $\hat{\alpha}$s and $\hat{\beta}$s come from the specification-adjusted production function estimates above. Lab capabilities, Lab TFPC, is simply the product of Lab TFPQ and Lab TFPZ.

As we are able to ensure that inputs and product specifications are identical across firms, we believe that the lab measures contain the least measurement error and come closest to reflecting firms' true productivity.[11] Thus, we treat them as benchmarks with which to assess our survey measures.

To summarize, the surveys provide two measures of quantity productivity (ϕ_u,ϕ_a), two of quality productivity (ζ_u,ζ_a), two of capability ($\zeta_u\phi_u$,$\zeta_a\phi_a$), and TFPR. The controlled lab provides three benchmarks: lab quantity productivity (Lab TFPQ), lab quality productivity (Lab TFPZ), and lab capabilities (Lab TFPC).

For each firm, we have one survey-based measure for each survey round. To reduce noise, we take firm-level averages over all post-baseline rounds and present all the productivity measures relative to the mean.[12] The Online Appendix provides further details on implementation and the production function estimates.

[9]This approach mirrors the price index literature with equation (2) acting as a hedonic regression that quality adjusts quantities before estimating the production function.

[10]Rug revenues and the value of k come from direct survey questions. Labor expenditures equal wages paid to employees and the take home pay of weaver-owners. Values are adjusted using the monthly CPI.

[11]Since the loom, specifications, and inputs are identical for all firms in the lab, we do not need to specification adjust to compare across firms.

[12]The experiment in Atkin, Khandelwal and Osman (2017) showed that inducing firms to export raised productivity. To ensure that we are not combining estimates for the same firm pre and post treatment, we only include the post treatment rounds.

4 Comparing Productivity Measures

We now explore the relationship between the various productivity measures and draw conclusions for practitioners working with less-rich datasets. We also discuss the dispersion in productivity across firms implied by each measure, a key moment of interest in the productivity literature.

Result 1: Importance of Adjusting for Product Specifications. Comparing unadjusted TFPQ across firms is challenging when specifications vary substantially.

Figure 1 plots both unadjusted and specification-adjusted TFPQ against Lab TFPQ—the measure we believe is closest to true quantity productivity.[13]

Consistent with the claim above, although the slope is positive (β = 0.13), unadjusted TFPQ only weakly correlates with Lab TFPQ ($corr$=0.02). Specification-adjusted TFPQ has a steeper slope and a stronger correlation with Lab TFPQ (β = 0.51,$corr$=0.14). This shows the value of finer product-category controls for accurately measuring quantity productivity.

Result 2: Quantity versus Quality Productivity. In this industry, as in many others, consumers place substantial value on quality. Our prior is that firms that are able to produce high quality are highly skilled, and so can also produce products with a given set of specifications more quickly. If there is a strong positive correlation between the two, then quality-productivity measures may do a satisfactory job at capturing a firm's broader capabilities.

Figure 2 shows two plots. The first reveals a strong negative relationship between unadjusted TFPQ and unadjusted TFPZ (black). Thus, the in the absence of specification controls, quantity and quality productivity are *negatively* correlated: firms that make lower quality rugs produce more quickly. However, and further showing the importance of adjusting for specifications, this relationship *flips* when we adjust for specifications (the second plot in gray, specification-adjusted TFPQ against specification-adjusted TFPZ). Consistent with our prior, quantity and quality productivity are positively related. More capable firms take longer to manufacture rugs only because they typically make varieties with more demanding specifications.

Thus, and as we show more directly below, in the absence of specification controls, TFPQ may be a misleading measure of broad capabilities given the strong negative correlation between unadjusted quantity and quality productivity.

Result 3: TFPR as a Proxy for TFPC. If capabilities are multidimensional, and consumers value quality, TFPR may be preferable to TFPQ-based measures since higher prices and revenues may

[13]The figures show both the line of best fit, the slope and significance of this line, the correlation coefficient, as well as a bin scatter of observations (each dot reflects about 10 firms). The online appendix reports a correlation matrix for the various measures.

capture the ability to produce high quality. Figure 3 explores this claim by comparing several of our productivity measures to Lab TFPC, the capability measure that combines quality and quantity productivity from the lab.

Consistent with the discussion above, unadjusted TFPQ is a misleading measure of capability: it is negatively correlated with Lab TFPC (black diamond). However, TFPR (grey circle) does indeed mitigate this measurement issue since it is positively correlated with Lab TFPC. Although the relationship is weak, this reversal of slope relative to unadjusted TFPQ reveals that TFPR may be a more suitable proxy for a firm's capability than TFPQ if product specifications are unavailable. Specification-adjusted TFPQ (black triangle) is more strongly positively correlated with Lab TFPC. As shown in Result 1, it more accurately captures quantity productivity, and as shown in Result 2, quantity and quality productivity are positively correlated after specification-adjusting. Reassuringly, specification-adjusted TFPC (grey square), which combines specification-adjusted TFPQ and TFPZ, has the strongest positive relationship with Lab TFPC.

Result 4: Unadjusted TFPQ Overstates Dispersion more than Specification-Adjusted TFPQ. Table 1 provides 90-10 ratios for the various productivity measures (figures 4-7 plot the distributions of TFPQ, TFPZ, TFPC and TFPR respectively). Dispersion in Lab TFPQ is over three times smaller than unadjusted TFPQ. Adjusting for specifications closes about half this gap. This suggests that dispersion in standard datasets may partially reflect product differentiation rather than differences in underlying productivity.

Result 5: TFPZ Dispersion is Large. Table 1 reveals large dispersion in quality productivity. The 90-10 ratio in Lab TFPZ is 2.2. From the surveys, the unadjusted and adjusted TFPZ ratios are 2.5 and 1.5, respectively. This suggests that even within a very narrowly defined product category, there is large quality variation across firms.

Result 6: TFPC is More Dispersed than TFPQ and TFPZ. Capabilities are even more dispersed than either quantity or quality productivity. The 90-10 ratio for Lab TFPC is larger than that for Lab TFPQ and Lab TFPZ (similarly for specification-adjusted TFPC). An implication of the fact that quantity and quality productivity are positively correlated, this result suggests that the broad capabilities of firms may be more dispersed than a single dimension of productivity. To our knowledge, this is the first attempt to document dispersion in capabilities through direct measurement.

5 Concluding Remarks

We close with a summary of this measurement exercise. First, standard TFPQ performs poorly at measuring quantity productivity. Using product specifications to make apples-to-apples comparisons substantially raises the correlation with the lab benchmarks and halves the gap in measured productivity dispersion between survey and lab measures. Second, firms differ substantially along a second dimension of productivity—their ability to produce high-quality products. Finally, if researchers are interested in broader capabilities of firms, TFPR—for all its imperfections—may be a better proxy than (unadjusted) TFPQ. TFPQ is likely to perform particularly poorly in settings like ours where more capable firms make products with more demanding specifications that take longer to manufacture. But, tailored surveys that collect product specifications and direct measures of quality may be the best path to understand productivity dispersion across firms.

References

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 132(2): 551–615.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains From Variety." *The Quarterly Journal of Economics*, 121(2): 541–585.

de Loecker, Jan, and Pinelopi Goldberg. 2014. "Firm Performance in a Global Market." *The Annual Review of Economics*, 6(1): 201–227.

Hallak, Juan Carlos, and Jagadeesh Sivadasan. 2013. "Product and process productivity: Implications for quality choice and conditional exporter premia." *Journal of International Economics*, 91(1): 53 – 67.

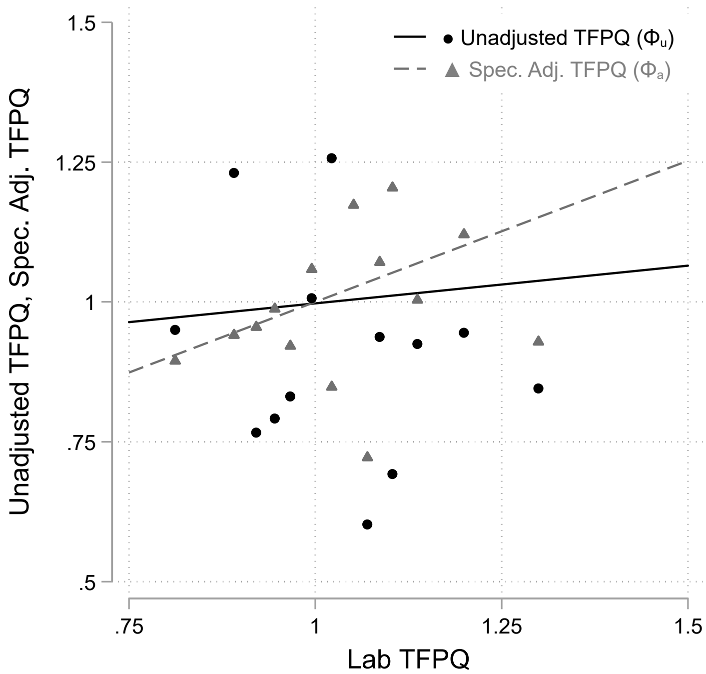
Syverson, Chad. 2011. "What Determines Productivity?" *Journal of Economic Literature*, 49(2): 326–65.

Table 1: Productivity Dispersion (90-10 Ratios)

Lab TFPQ	1.3	Lab TFPZ	2.2
Lab TFPC	2.3	TFPR	2.7
Unadj TFPQ (ϕ_u)	4.7	Adj TFPQ (ϕ_a)	3.1
Unadj TFPZ (ζ_u)	2.5	Adj TFPZ (ζ_a)	1.5
Unadj TFPC ($\zeta_u\phi_u$)	4.3	Adj TFPC ($\zeta_a\phi_a$)	3.5

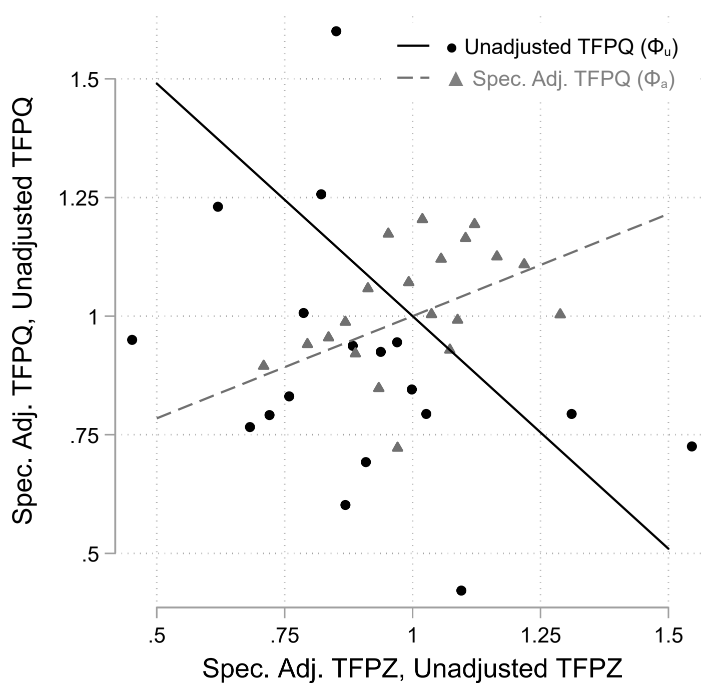
Notes: Table reports 90-10 ratios for productivity measures.

Figure 1: Adjusting for Product Specifications



Φ_u: β=0.13 (se=0.66,N=186) corr=0.02 Φ_a: β=0.51 (se=0.27,N=186) corr=0.14

Figure 2: Quantity versus Quality Productivity



Φ_u: β=-0.98 (se=0.16,N=209) corr=-0.40 Φ_a: β=0.43 (se=0.19,N=209) corr=0.15

Figure 3: TFPR as Substitute for TFPC

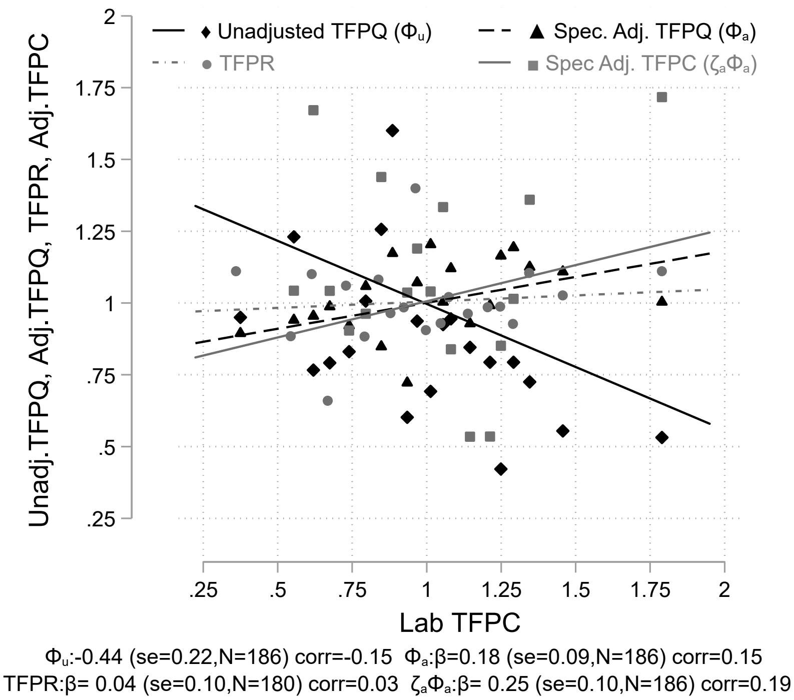


Figure 4: Distribution of TFPQ

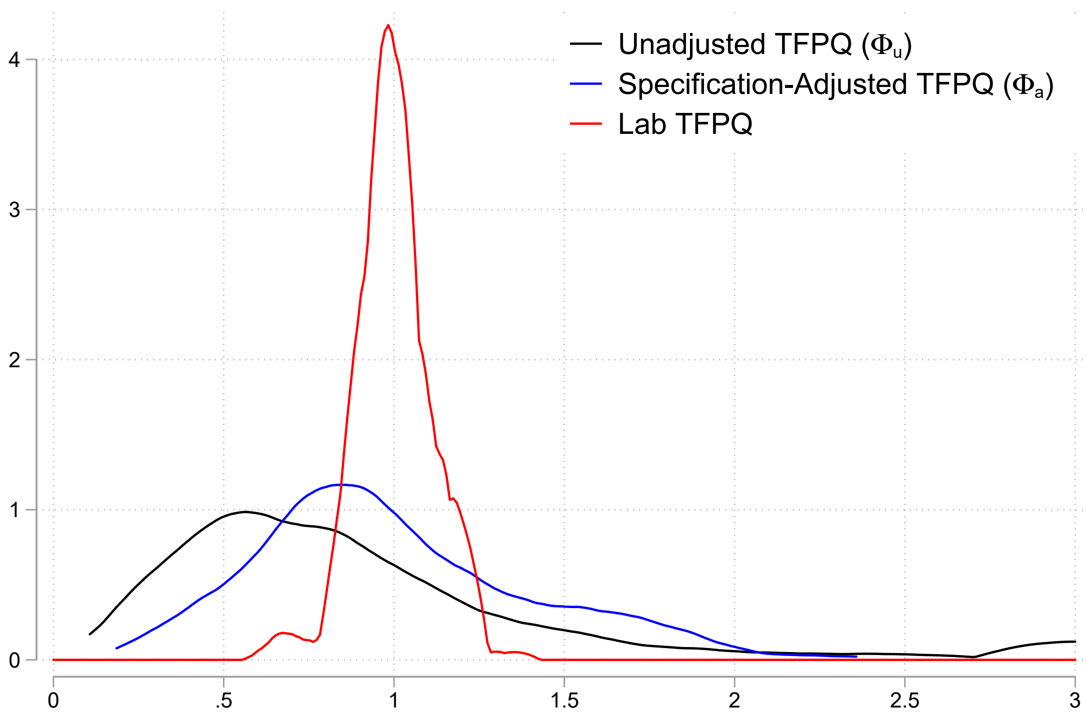


Figure 5: Distribution of TFPZ

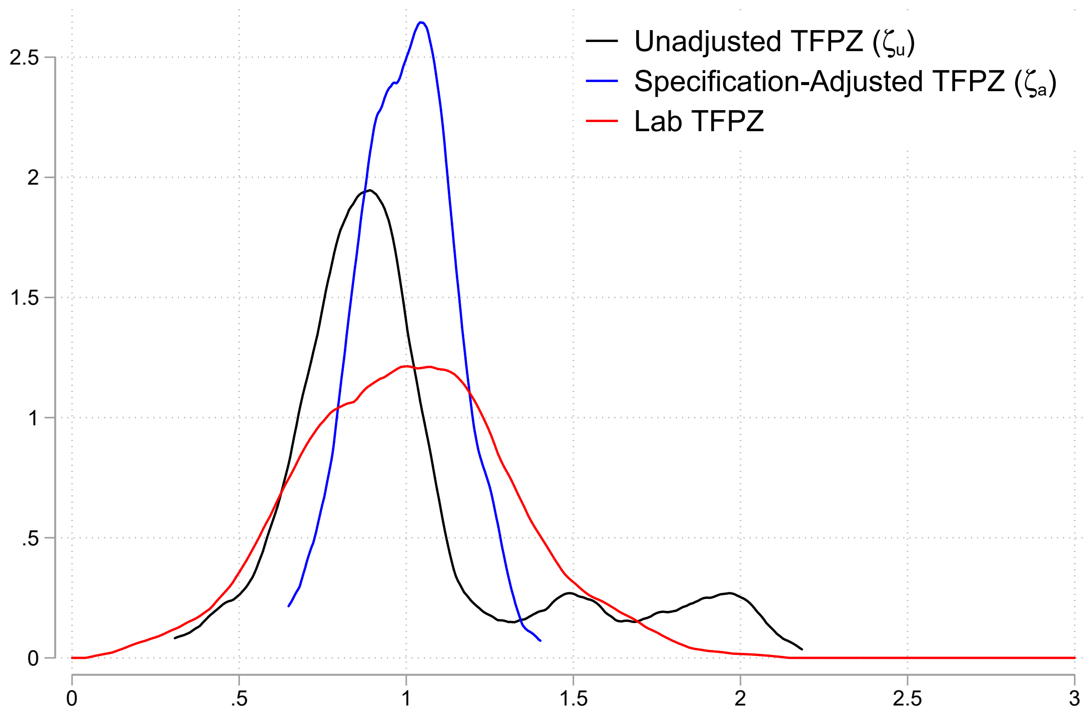


Figure 6: Distribution of TFPC

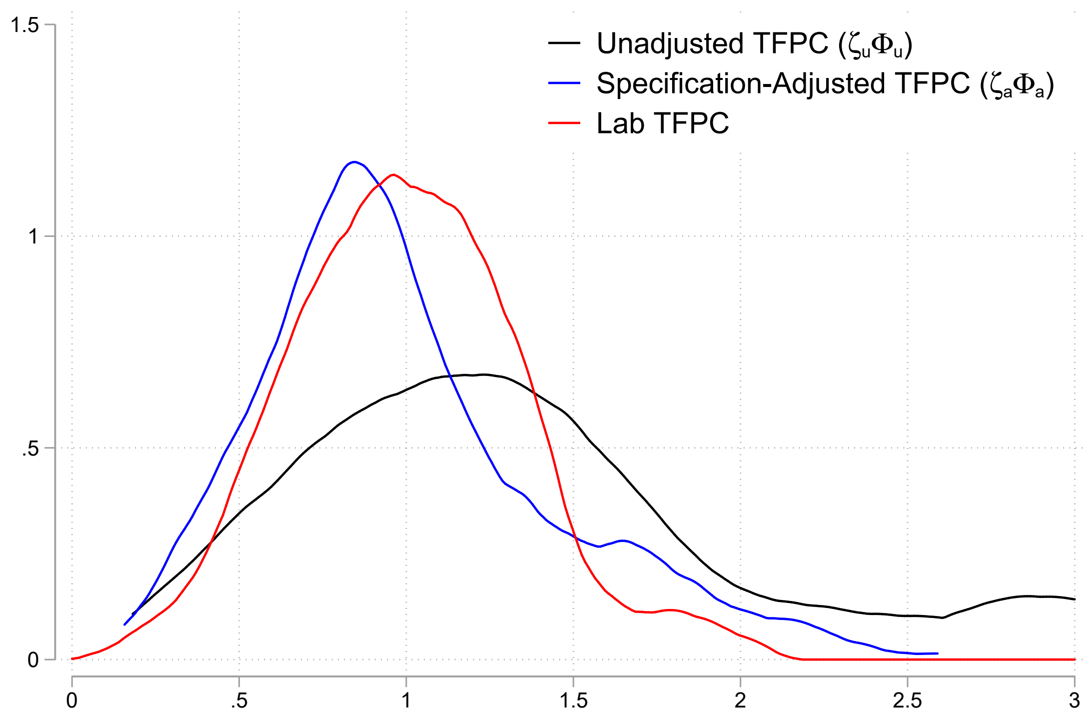
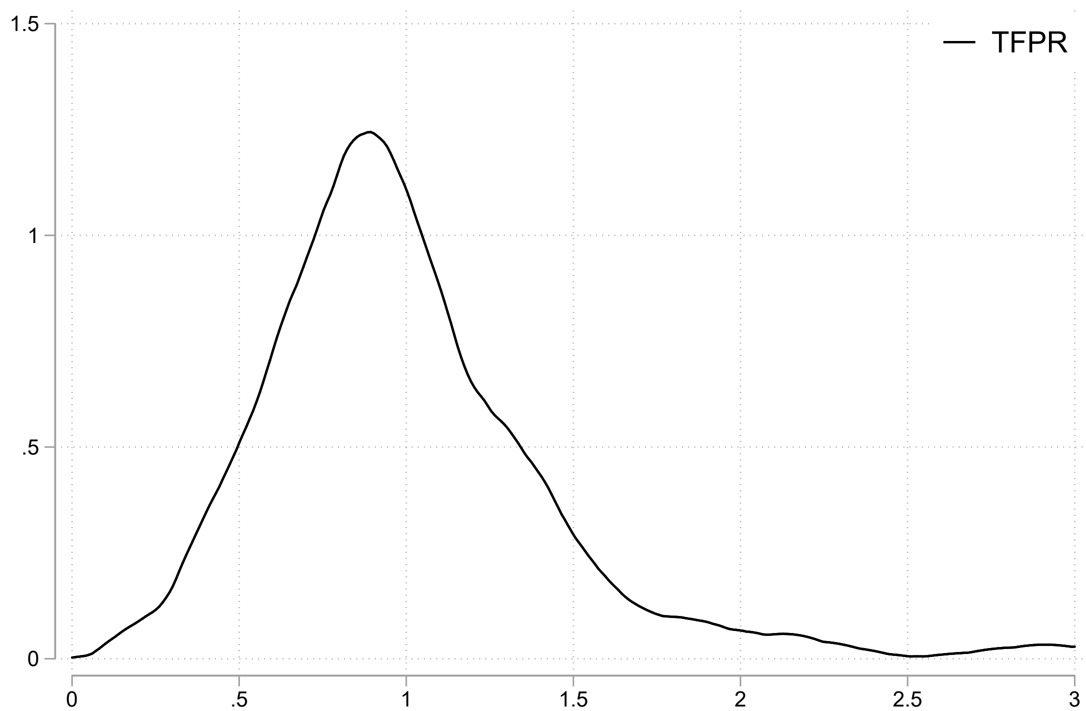


Figure 7: Distribution of TFPR



Appendix

A.1 Summary Statistics

We collected data for 219 rug producing firms in Fowa, Egypt between July 2011 and June 2014. We administered six rounds of surveys that captured information on rugs produced in the prior month including the rug specifications, prices and quantities of all inputs and outputs, labor hours spent on production and preparation activities. We also hired an independent quality assessor (a highly-skilled rug producer) who graded the rugs being produced at the time of the survey along eleven quality metrics (grading on a 1 to 5 scale, with 5 being the highest quality). Table A.1 provides the means and standard deviations of key variables used in our estimations.

After the last survey round, we set up a controlled lab in a rented space where all firms were asked to send their main rug producer to produce a rug with identical specifications using the same material inputs and capital equipment that we provided. The rug producer was paid a flat fee for his time. We recorded the rug's final dimensions and the time taken to weave it. We also sent the rugs to be scored anonymously by both our quality assessor and a local professor of handicraft science. We use the average score for each quality metric in this paper. Table A.1 also reports the mean and standard deviations of the quality lab measures. Atkin, Khandelwal and Osman (2017) provides further details on the surveys and the sample.

Table A.2 reports the association between the rug specifications and the price of the rugs, overall output and total revenue of the firm during the month prior to the survey. The coefficients have signs consistent with our priors and the high R-squared suggests that specifications can explain much of the variation in these variables.

A.2 Survey-Based Productivity Measures

A.2.1 Quantity Production Functions

Our production function estimation follows Atkin, Khandelwal and Osman (2017). The first set of production function estimates do not control for rug specifications and hence provide our *unadjusted TFPQ* estimates. We estimate the following Cobb-Douglas production function:[14]

$$x_{it} = \phi_{u,it} l_{it}^{\alpha_l} k_{it}^{\alpha_k} e^{\epsilon_{it}} \tag{4}$$

[14] We assume that output is Leontief in materials and therefore materials do not enter into the estimation.

where x_{it} is the output (in m^2) of firm i in period t, l_{it} is total labor hours, k_{it} is the number of active looms, and $\phi_{u,it}$ is the firm's unadjusted TFPQ. The error term captures unanticipated shocks as well as omitted variables (the specifications of the rugs produced). To estimate the parameters of the production function, we take logs to obtain

$$\ln x_{it} = \ln \phi_{u,it} + \alpha_l^u \ln l_{it} + \alpha_k^u \ln k_{it} + \epsilon_{it} \tag{5}$$

The second set of production function estimates controls for rug specifications and provide our *specification-adjusted TFPQ* estimate. We estimate

$$\ln x_{it} = \ln \phi_{a,it} + \alpha_l^a \ln l_{it} + \alpha_k^a \ln k_{it} + \ln \lambda_{it}^{'} \gamma + \epsilon_{it} \tag{6}$$

where $\phi_{a,it}$ is the firm's specification-adjusted TFPQ and the vectorλ_{it} includes six rug specifications: rug difficulty, thread count, thread type, number of colors, market segment, and narrow product type (where we include dummies for each value of the latter two categorical variables).[15] The error term now only captures unanticipated shocks and measurement error.

We estimate TFPQ via OLS and a control function. For the OLS regressions, we estimate (5) by regressing log of output on labor and capital. For (6), we add the six specifications to the regression. We estimate the production functions using the full set of duble firms in our sample of post-treatment rounds.[16] Standard errors are clustered by firm. We report the estimates in columns 1 and 2 of Table A.3 below.

In the control function approach (Olley and Pakes (1996)) we assume capital is subject to adjustment costs, labor is a flexible input, and we use warp thread quantity as the proxy. We estimate the production functions using the one-step approach proposed by Wooldridge (2009), with l_{it-1} as the instrument for l_{it}, and cluster standard errors by firm. We report these estimates in columns 3 and 4 in Table A.3 below.

Unadjusted and specification-adjusted TFPQ are constructed from exponentiating the residuals of these production functions and then averaging across rounds for each firm.

[15]As discussed in Atkin, Khandelwal and Osman (2017), we have two samples of firms that we pool over in this production function estimation. For the firms in the first sample, we did not record the market segment or rug difficulty. We replace these missing values with the corresponding values from the subsequent survey round.

[16]This differs from Atkin, Khandelwal and Osman (2017) where we estimate the production function only on the sample of control firms to avoid having to take a stance on the Markov process governing productivity evolution over time for the treatment firms. In this paper, since we are simply interested in cross-sectional comparisons, we only focus on the post-treatment sample where export status is not changing and estimate the production function over all firms.

A.2.2 Quality Production Functions

Quality productivity, TFPZ, is estimated as follows. As noted in the text, we begin by obtaining the consumers' valuations for quality implied by the following demand curve:

$$\ln x_{it} = (\sigma - 1) \sum_j \theta_j \ln q_{j,it} - \sigma \ln p_{it} + c_{it} \tag{7}$$

where q_js are the eleven quality metrics, p is the price that firm i receives for its rug produced at the time of the survey, and c is a common price index. Using an estimate of $\sigma = 2.74$ from Broda and Weinstein (2006), we can re-write (7) as an estimating equation:

$$(\ln x_{it} + 2.74 \ln p_{it}) \,/\, (2.74 - 1) = \kappa + \sum_j \theta_j \ln q_{j,it} + \nu_{it} \tag{8}$$

where κ is a constant and ν is measurement error. The estimates of the θ_js are reported in Table A.4.

With the estimates of θ in hand, we formulate the production function for producing consumers' valued quality, $\Pi_j q_j^{\hat{\theta}_j}$, with the same functional form as the quantity production function in (5):

$$\ln \left(\Pi_j q_{j,it}^{\hat{\theta}_j} \right) = \ln \zeta_{u,it} + \beta_l^u \ln l_{it} + \beta_k^u \ln k_{it} + \epsilon_{it} \tag{9}$$

As before, we can estimate (9) via OLS or a control function to obtain *unadjusted TFPZ*. The results are reported in Table A.5.

Analogously to specification-adjusted TFPQ, we can recover *specification-adjusted TFPZ* by controlling for specifications in the quantity production production:

$$\ln \left(\Pi_j q_{j,it}^{\hat{\theta}_j} \right) = \ln \zeta_{a,it} + \beta_l^a \ln l_{it} + \beta_k^a \ln k_{it} + \ln \lambda_{it}^{'} \delta + \epsilon_{it} \tag{10}$$

The results of estimating (10) via OLS and a control function are reported in Table A.5.

Unadjusted and specification-adjusted TFPZ are constructed from exponentiating the residuals of these production functions and then averaging across rounds for each firm.

A.2.3 Capabilities Production Functions

For unadjusted firm capabilities, which we term *unadjusted TFPC*, we multiply output by the quality aggregator to formulate a combined production function for $x_{it} \Pi_j q_{j,it}^{\hat{\theta}_j}$, the combination of quantity and quality that consumers value in their utility function.

$$\ln\left(x_{it}\Pi_j q_{j,it}^{\hat{\theta}_j}\right) = \ln\zeta_{u,it} + \ln\phi_{u,it} + (\alpha_l^u + \beta_l^u)\ln l_{it} + (\alpha_k^u + \beta_k^u)\ln k_{it} + \epsilon_{it} \tag{11}$$

As before, we estimate (11) via OLS and a control function, and report the results in Table A.6. The structure of the production function implies that the coefficients of the capabilities production function equal the sum of the coefficients from the quantity and quality production functions (e.g., the sum of the labor coefficient in column 1 of Table A.3 and the labor coefficient in column 1 of Table A.5.[17]). Unadjusted TPFC is the product $\zeta_u\phi_u$.

Similarly, we can estimate *specification-adjusted TFPC* from the following production function:

$$\ln\left(x_{it}\Pi_j q_{j,it}^{\theta_j}\right) = \ln\zeta_{a,it} + \ln\phi_{a,it} + (\alpha_l^a + \beta_l^a)\ln l_{it} + (\alpha_k^a + \beta_k^a)\ln k_{it} + \ln\lambda_{it}^{'}(\gamma+\delta) + \epsilon_{it} \tag{12}$$

with the results reported in A.6. Specification-adjusted TPFC is the product $\zeta_a\phi_a$.

A.2.4 Revenue Production Functions

We estimate a revenue production function using the following specification:

$$\ln r_{it} = \ln TFPR_{it} + \eta_l \ln w_{it} + \eta_k \ln rk_{it} + \epsilon_{it} \tag{13}$$

where r_{it} is the revenue of the firm, w_{it} is the wage bill, and rk_{it} is the value of the capital stock. We estimate (13) via OLS and a control function and report the results in A.7. (Note that we do not control for specifications in these regressions). TFPR is constructed from exponentiating the residual of this production function and then averaging across rounds for each firm.

A.3 Description of Appendix Figures and Tables

- Table A.1 provides summary statistics for the variables used to estimate the production functions.
- Table A.2 estimates the relationship between the rug specifications and price, output and revenue.
- Table A.3 reports the coefficients from the quantity production function.

[17]Due to missing observations, the coefficients do not line up exactly. The paper uses the TFPC estimate that comes from the product of the individual ζ_u and ϕ_u estimates.

- Table A.4 reports the θs coefficients from the demand estimation.
- Table A.5 reports the coefficients from the quality production function.
- Table A.6 reports the coefficients from the capabilities production function.
- Table A.8 is the correlation matrix for the measures used in the paper estimated using OLS.
- Table A.9 is the correlation matrix for the measures used but estimated using a control function approach.
- Table A.10 shows the correlation matrix including both OLS and control function values.

References

Atkin, David, Amit K. Khandelwal, and Adam Osman. 2017. "Exporting and Firm Performance: Evidence from a Randomized Experiment." *The Quarterly Journal of Economics*, 132(2): 551–615.

Broda, Christian, and David E. Weinstein. 2006. "Globalization and the Gains From Variety." *The Quarterly Journal of Economics*, 121(2): 541–585.

Olley, G Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, 64(6): 1263–97.

Wooldridge, Jeffrey M. 2009. "On estimating firm-level production functions using proxy variables to control for unobservables." *Economics Letters*, 104(3): 112–114.

A.4 Appendix Tables

Table A.1: Summary Statistics

	Mean	Standard Deviation	Observations
Output (Square Meters)	59.43	(75.04)	900
Labor Hours	5.55	(0.29)	900
Capital (Looms)	0.08	(0.27)	912
(log) Thread Quantity	7.46	(0.28)	913
Difficulty Control	3.23	(0.83)	926
(log) Number of Colors	1.72	(0.99)	914
Mid-Market Segment=1	0.23	(0.42)	923
Low-Market Segment=1	0.42	(0.49)	923
Price (EGP/Square Meter)	29.29	(47.27)	913
Survey Quality: Packedness	3.25	(0.86)	913
Survey Quality: Corners	3.14	(0.85)	913
Survey Quality: Waviness	3.15	(0.84)	913
Survey Quality: Weight	3.22	(0.84)	913
Survey Quality: Touch	3.19	(0.49)	913
Survey Quality: Warp Thread Tightness	3.18	(0.81)	913
Survey Quality: Firmness	3.02	(0.56)	913
Survey Quality: Design Accuracy	3.32	(0.86)	913
Survey Quality: Ward Thread Packedness	3.19	(0.83)	913
Survey Quality: Inputs	3.20	(0.87)	913
Survey Quality: Loom	2.04	(0.24)	913
Lab Quality: Packedness	3.34	(0.63)	187
Lab Quality: Corners	3.29	(0.63)	187
Lab Quality: Waviness	3.28	(0.60)	187
Lab Quality: Weight	3.60	(0.83)	187
Lab Quality: Touch	3.29	(0.50)	187
Lab Quality: Warp Thread Tightness	2.95	(0.66)	187
Lab Quality: Firmness	3.24	(0.65)	187
Lab Quality: Design Accuracy	3.46	(0.62)	187
Lab Quality: Ward Thread Packedness	3.27	(0.68)	187
Lab Quality: Inputs	4.00	(0.00)	187
Lab Quality: Loom	2.00	(0.00)	187

Notes: Table reports summary statistics of the variables used to estimate the production functions. "Quality" denotes the 11 quality metrics. "Lab" denotes the quality metrics from the controlled lab, which are averaged over grades given by the quality assessor and professor of handicraft science. "EGP" denotes Egyptian pounds (which was around 6.31 pounds to one USD over the sample period). See Atkin et al (2017) for more details about the sample and variables.

Table A.2: Outcomes and Specifications

	(1) Price	(2) Output	(3) Revenue
(log) Thread Quantity	0.11	-0.01	0.11
	(0.14)	(0.12)	(0.11)
Difficulty Control	0.13***	-0.06*	0.07**
	(0.03)	(0.03)	(0.03)
(log) Number of Colors	-0.02	-0.05*	-0.06**
	(0.03)	(0.03)	(0.03)
Low-Market Segment=1	-0.84***	0.52***	-0.30***
	(0.08)	(0.07)	(0.07)
Mid-Market Segment=1	-0.60***	0.31***	-0.26***
	(0.08)	(0.08)	(0.06)
Product Type Dummies (6 Categories)	Yes	Yes	Yes
Thread Type Dummies (6 Categories)	Yes	Yes	Yes
r2	.536	.454	.117
N	825	890	818

Notes: Table reports the results of estimating the log price, log output and log revenue on the six specifications. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.3: Quantity Production Function

	(1) Unadjusted (OLS)	(2) Adjusted (OLS)	(3) Unadjusted (CF)	(4) Adjusted (CF)
Labor	0.61***	0.65***	1.41**	1.31***
	(0.11)	(0.09)	(0.70)	(0.51)
Capital Inputs	0.19*	0.24**	0.41**	0.24**
	(0.11)	(0.10)	(0.19)	(0.12)
(log) Thread Quantity		-0.02		-0.28*
		(0.11)		(0.17)
Difficulty Control		-0.06**		-0.12***
		(0.03)		(0.04)
(log) Number of Colors		-0.05*		-0.07**
		(0.03)		(0.03)
Low-Market Segment=1		0.56***		0.55***
		(0.07)		(0.08)
Mid-Market Segment=1		0.37***		0.34***
		(0.07)		(0.08)
Product Type Dummies (6 Categories)	No	Yes	No	Yes
Thread Type Dummies (6 Categories)	No	Yes	No	Yes
r2	.046	.506	.000	.508
N	899	889	595	591

Notes: Table reports the results of estimating the quantity production function. Columns 1 and 3 estimate the unadjusted production function. Columns 2 and 4 estimate the specification-adjusted production function. Columns 1-2 estimate via OLS and columns 3-4 estimate via a control function. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.4: Consumers' Valuation of Quality ($\theta' s$)

	(1)
	Consumer Quality Valuation
Packedness	0.18
	(0.26)
Corners	-0.10
	(0.26)
Waviness	-0.09
	(0.26)
Weight	-0.10
	(0.22)
Touch	0.15
	(0.27)
Warp Thread Tightness	0.87***
	(0.25)
Firmness	-0.31
	(0.33)
Design Accuracy	0.76***
	(0.20)
Ward Thread Packedness	0.51**
	(0.24)
Inputs	-0.09
	(0.23)
Loom	-0.70*
	(0.41)
r2	.168
N	892

Notes: Table reports the results of estimating the demand curve to obtain consumers' valuation of quality, θ_js. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.5: Quality Production Function

	(1) Unadjusted (OLS)	(2) Adjusted (OLS)	(3) Unadjusted (CF)	(4) Adjusted (CF)
Labor	0.07	-0.01	-0.18	-0.05
	(0.04)	(0.03)	(0.34)	(0.12)
Capital Inputs	0.01	0.08**	-0.02	0.09*
	(0.05)	(0.03)	(0.08)	(0.05)
(log) Thread Quantity		0.02		0.08
		(0.03)		(0.06)
Difficulty Control		0.37***		0.35***
		(0.01)		(0.02)
(log) Number of Colors		0.02**		0.02
		(0.01)		(0.01)
Low-Market Segment=1		-0.07***		-0.10***
		(0.02)		(0.03)
Mid-Market Segment=1		-0.07***		-0.09***
		(0.03)		(0.03)
Product Type Dummies (6 Categories)	No	Yes	No	Yes
Thread Type Dummies (6 Categories)	No	Yes	No	Yes
r2	.002	.672	.052	.742
N	891	882	589	585

Notes: Table reports the results of estimating the quality production function. Columns 1 and 3 estimate the unadjusted production function. Columns 2 and 4 estimate the specification-adjusted production function. Columns 1-2 estimate via OLS and columns 3-4 estimate via a control function. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.6: Capabilities Production Function

	(1) Unadjusted (OLS)	(2) Adjusted (OLS)	(3) Unadjusted (CF)	(4) Adjusted (CF)
Labor	0.67***	0.63***	1.32**	1.12***
	(0.10)	(0.09)	(0.65)	(0.35)
Capital Inputs	0.20*	0.32***	0.35**	0.28**
	(0.10)	(0.10)	(0.15)	(0.14)
(log) Thread Quantity		0.00		-0.19
		(0.11)		(0.19)
Difficulty Control		0.30***		0.23***
		(0.03)		(0.04)
(log) Number of Colors		-0.02		-0.05
		(0.03)		(0.03)
Low-Market Segment=1		0.47***		0.43***
		(0.07)		(0.08)
Mid-Market Segment=1		0.29***		0.24***
		(0.08)		(0.08)
Product Type Dummies (6 Categories)	No	Yes	No	Yes
Thread Type Dummies (6 Categories)	No	Yes	No	Yes
r2	.062	.341	.005	.279
N	891	882	589	585

Notes: Table reports the results of estimating the capability production function. Columns 1 and 3 estimate the unadjusted production function. Columns 2 and 4 estimate the specification-adjusted production function. Columns 1-2 estimate via OLS and columns 3-4 estimate via a control function. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.7: Revenue Production Function

	(1)
	Log Revenue
Wage Bill	0.46***
	(0.10)
Value of Captial Stock	0.08**
	(0.04)
r2	.070
N	794

Notes: Table reports the results of estimating the revenue production function. Column 1 estimates via OLS and column 2 estimates via a control function. Standard errors clustered at the firm level in parentheses. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.8: Correlation Matrix (OLS)

	Lab TFPQ	Lab TFPC	Lab TFPZ	Unadj. TFPQ	Adj. TFPQ	TFPR	Unadj. TFPZ	Adj. TFPZ	Unadj. TFPC
Lab TFPC	0.40***	1.00							
Lab TFPZ	0.07	0.94***	1.00						
Unadj. TFPQ	0.02	-0.15**	-0.15**	1.00					
Adj. TFPQ	0.14*	0.15**	0.11	0.42***	1.00				
TFPR	-0.02	0.03	0.03	0.10	0.28***	1.00			
Unadj. TFPZ	-0.05	0.45***	0.50***	-0.40***	0.01	0.07	1.00		
Adj. TFPZ	-0.17**	0.14*	0.22***	0.08	0.15**	-0.05	0.52***	1.00	
Unadj. TFPC	0.03	0.10	0.11	0.78***	0.71***	0.20***	0.07	0.40***	1.00
Adj. TFPC	0.07	0.19***	0.19**	0.38***	0.94***	0.24***	0.18***	0.44***	0.77***

Notes: Table reports the correlation between the variable at the top of each column with the variable in the associated row. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.9: Correlation Matrix (Control Function)

	Lab TFPQ	Lab TFPC	Lab TFPZ	Unadj. TFPQ	Adj. TFPQ	TFPR (CF)	Unadj. TFPZ	Adj. TFPZ	Unadj. TFPC
Lab TFPC	0.40***	1.00							
Lab TFPZ	0.07	0.94***	1.00						
Unadj. TFPQ (CF)	0.03	-0.15**	-0.16**	1.00					
Adj. TFPQ (CF)	0.16**	0.18**	0.14*	0.45***	1.00				
TFPR (CF)	0.01	0.04	0.02	0.07	0.22***	1.00			
Unadj. TFPZ (CF)	-0.04	0.45***	0.50***	-0.40***	0.03	0.11	1.00		
Adj. TFPZ (CF)	-0.17**	0.13*	0.21***	0.18***	0.16**	-0.08	0.48***	1.00	
Unadj. TFPC (CF)	0.04	0.09	0.09	0.81***	0.70***	0.17**	0.02	0.44***	1.00
Adj. TFPC (CF)	0.10	0.22***	0.20***	0.45***	0.94***	0.19***	0.17**	0.43***	0.77***

Notes: Table reports the correlation between the variable at the top of each column with the variable in the associated row. Significance: * 0.10, ** 0.05, *** 0.01.

Table A.10: Correlation Matrix (OLS and Control Function)

	Lab TFPQ	Lab TFPC	Lab TFPZ	Unadj. TFPQ	Adj. TFPQ	Unadj. TFPQ (CF)	Adj. TFPQ (CF)	TFPR	TFPR (CF)	Unadj. TFPZ	Adj. TFPZ	Unadj. TFPZ (CF)	Adj. TFPZ (CF)	Unadj. TFPC	Adj. TFPC	Unadj. TFPC (CF)
Lab TFPC	0.40***	1.00														
Lab TFPZ	0.07	0.94***	1.00													
Unadj. TFPQ	0.02	-0.15**	-0.15**	1.00												
Adj. TFPQ	0.14*	0.15**	0.11	0.42***	1.00											
Unadj. TFPQ (CF)	0.03	-0.15**	-0.16**	0.98***	0.43***	1.00										
Adj. TFPQ (CF)	0.16**	0.18**	0.14*	0.42***	0.94***	0.45***	1.00									
TFPR	-0.02	0.03	0.03	0.10	0.28***	0.08	0.22***	1.00								
TFPR (CF)	0.01	0.04	0.02	0.08	0.27***	0.07	0.22***	0.97***	1.00							
Unadj. TFPZ	-0.05	0.45***	0.50***	-0.40***	0.01	-0.40***	0.04	0.07	0.09	1.00						
Adj. TFPZ	-0.17**	0.14*	0.22***	0.08	0.15**	0.09	0.17**	-0.05	-0.06	0.52***	1.00					
Unadj. TFPZ (CF)	-0.04	0.45***	0.50***	-0.39***	0.01	-0.40***	0.03	0.08	0.11	1.00***	0.51***	1.00				
Adj. TFPZ (CF)	-0.17**	0.13*	0.21***	0.17**	0.15**	0.18***	0.16**	-0.07	-0.08	0.49***	0.98***	0.48***	1.00			
Unadj. TFPC	0.03	0.10	0.11	0.78***	0.71***	0.79***	0.69***	0.20***	0.20***	0.07	0.40***	0.07	0.46***	1.00		
Adj. TFPC	0.07	0.19***	0.19**	0.38***	0.94***	0.40***	0.89***	0.24***	0.24***	0.18***	0.44***	0.18***	0.43***	0.77***	1.00	
Unadj. TFPC (CF)	0.04	0.09	0.09	0.77***	0.70***	0.81***	0.70***	0.17**	0.17**	0.04	0.40***	0.02	0.44***	0.98***	0.75***	1.00
Adj. TFPC (CF)	0.10	0.22***	0.20***	0.42***	0.87***	0.45***	0.94***	0.18***	0.19***	0.19***	0.42***	0.17**	0.43***	0.76***	0.94***	0.77***

Notes: Table reports the correlation between the variable at the top of each column with the variable in the associated row. Significance: * 0.10, ** 0.05, *** 0.01.