THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. 132 May 2017 Issue 2

EXPORTING AND FIRM PERFORMANCE: EVIDENCE FROM A RANDOMIZED EXPERIMENT*

DAVID ATKIN
AMIT K. KHANDELWAL
ADAM OSMAN

We conduct a randomized experiment that generates exogenous variation in the access to foreign markets for rug producers in Egypt. Combined with detailed survey data, we causally identify the impact of exporting on firm performance. Treatment firms report 16–26% higher profits and exhibit large improvements in quality alongside reductions in output per hour relative to control firms. These findings do not simply reflect firms being offered higher margins to

*We thank Abdelrahman Nagy, Magdy El-Mezain, Atef Helmy, Salah El-Gazar, Aslam Ismail, Mabrook Abou Shaheen, Sherif Abou Shady and the firms who participated in our sample. We also thank the staff at Aid to Artisans and Creative Learning including Lisa Smith, Shereen Shirazy, Mary Cockram, Maud Mabika, and Bill Kruvant, as well as Ray Fisman and Christopher Udry for introducing us. We are grateful to Josephine Gantois for her excellent research assistance. We thank the editor (Pol Antràs) and five anonymous referees. We also thank Nicholas Bloom, Juan Carlos Hallak, David McKenzie, Nina Pavcnik, Jim Tybout, Eric Verhoogen, Christopher Woodruff, and multiple seminar participants for useful comments. 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. Yale IRB Approval #1001006247 (Atkin and Osman were at Yale University during the experiment), Columbia IRB Approval #AAAE9678.

© The Authors 2017. Published by Oxford University Press on behalf of President and Fellows of Harvard College. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com
The Quarterly Journal of Economics (2017), 551–615. doi:10.1093/qje/qjx002.
Advance Access publication on February 17, 2017.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

manufacture high-quality products that take longer to produce. Instead, we find evidence of learning-by-exporting whereby exporting improves technical efficiency. First, treatment firms have higher productivity and quality after controlling for rug specifications. Second, when asked to produce an identical domestic rug using the same inputs and same capital equipment, treatment firms produce higher quality rugs despite no difference in production time. Third, treatment firms exhibit learning curves over time. Finally, we document knowledge transfers with quality increasing most along the specific dimensions that the knowledge pertained to. *JEL Codes*: F10, F14, D24.

I. Introduction

There are large differences in productivity across countries (Hall and Jones 1999; Bloom and Van Reenen 2007). The belief that access to high-income markets can help firms in developing countries close this gap is one motivation behind the large resources now flowing to market access initiatives. For example, the World Trade Organization's Aid-for-Trade Initiative secured $48 billion in annual commitments to help developing countries overcome "trade-related constraints," and the last two decades have seen a tripling in the number of national export-promotion agencies that help domestic firms match with foreign buyers (Lederman, Olarreaga, and Payton 2010). Central to these programs achieving this goal is the belief that exporting improves the productivity of firms, a mechanism called learning-by-exporting (Clerides, Lach, and Tybout 1998; de Loecker 2007; Harrison and Rodriguez-Clare 2010).

In the presence of learning-by-exporting, trade generates efficiency gains which narrow this productivity gap and magnify the gains from trade relative to models without learning, such as those explored by Arkolakis, Costinot, and Rodriguez-Clare (2012) (e.g., see Alvarez, Buera, and Lucas 2013).

Despite the pervasiveness of these initiatives, there is still an ongoing debate as to whether exporting has a causal impact on measures of firm performance. Moreover, if performance does improve, it is unclear whether such improvements occur through learning-by-exporting—outward shifts in the production possibility frontier (PPF)—or simply through movements along the PPF. There are two central challenges in answering these questions. First, more productive firms select into exporting (for example, see Melitz 2003). This selection has plagued empirical attempts to identify the causal impact of exporting on firm performance

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

because what appears to be higher productivity among exporters may simply be self-selection. The second difficulty is that researchers typically lack detailed information required to isolate changes that occur within firms due to exporting. The literature commonly uses revenue-based total factor productivity (TFPR) measures which also reflect changes in markups, the product mix, and product quality (de Loecker and Goldberg 2014). This is problematic for identifying learning-by-exporting since the international trade literature suggests all three are likely to change with exporting.

While quantity-based total factor productivity measures (TFPQ) solve problems related to changing markups, standard data sets do not provide the level of detail required to account for changes in product specifications or quality.[1] Hence, if trade causes firms to change along these dimensions, measured improvements in quantity-based productivity measures may simply reflect movements along the PPF, rather than outward shifts of the PPF.

This article conducts a randomized controlled trial (RCT) on rug manufacturers in Egypt to examine how exporting affects profits and productivity. To our knowledge, this is the first attempt to generate exogenous firm-level variation in exporting. As explained in detail below, we achieved this through an intervention that reduced matching frictions between foreign buyers and a random subset of Egyptian firms. Using this experimental variation, we uncover if and how an economic primitive—firm productivity—responds to exporting.

The random assignment into exporting directly addresses the first of the two challenges detailed above: selection of firms into exporting. We provided a subset of firms with the opportunity to export handmade carpets to high-income markets. To provide this opportunity, we partnered with a U.S.-based nongovernmental organization (NGO) and an Egyptian intermediary to secure export orders from foreign buyers through trade fairs and direct marketing channels. With orders in hand, we surveyed a sample of several hundred small rug manufacturers located in Fowa, Egypt. A random subsample of these firms was provided with an

1. One solution is to restrict attention to homogeneous goods, such as concrete, block ice, or gasoline (e.g., Foster, Haltiwanger, and Syverson 2008). This is unappealing for the study of learning-by-exporting since there is likely to be less scope for learning and there are fewer trading frictions in homogeneous goods industries.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

initial opportunity to fill the orders by producing 110 m^2 of rugs (approximately 11 weeks of work). As in a standard buyer-seller relationship, firms were offered subsequent orders provided they were able to fulfill the initial orders to the satisfaction of the buyer and intermediary. Prior to our study, only a limited number of firms had ever knowingly exported their products. Hence, we interpret our experimental design as providing nonexporting firms with the opportunity to sell to high-income markets.

To address the second challenge in identifying the impact of exporting—measurement—we tracked performance measures through periodic surveys of both treatment firms (those who received the opportunity to export) and control firms (those who received no such opportunity). Our production-line level data allow us to record not just quantity data but also detailed specifications for the rugs being produced at the time of each survey round. These specifications include product categories within the flatweave-rug segment and attributes, such as the thread count, which a buyer chooses when the order is placed. This level of detail allows us to control for changes in the product mix due to exporting with much more accuracy than is possible in typical data sets (e.g., using HS-10 product codes in trade data sets). To further guarantee we are not conflating changes in productivity with changes in the product mix, at the end of the study all firms were paid to make an identical domestic-market rug using the same inputs and equipment. To analyze changes in product quality with exporting, we collect direct measures of product quality along 11 dimensions from a skilled quality assessor who visited each firm in each survey round. These quality measures capture a combination of both specifications and hard-to-codify attributes that depend on the technical skill of the firm, such as how flat the rug lies on the floor or how sharp the corners are. Finally, we collect data on information flows between buyers, the intermediary and producers that include transcripts of buyer feedback and the content of discussions between the intermediary and the producers. Taken together, these data allow us to address directly the measurement challenges noted above.

Thanks to the randomization procedure, the causal effects of exporting are identified by comparing mean outcomes between treatment and control firms. We find that the opportunity to export raises the overall performance of firms as measured by profits—treatment firms report 16–26% higher profits relative to control firms. The substantial increase in profits is perhaps not

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

surprising given that firms were provided with a positive demand shock, but is interesting given the more moderate profit impacts the literature has found when exploring supply-side interventions such as credit access (Banerjee 2013).

The primary focus of this article is to understand the mechanisms driving the profit increases. Despite increases in output prices and labor hours, we observe a decline in total output (m^2 of rugs produced) among treatment firms. These findings suggest that buyers from high-income countries demand higher-quality rugs that take longer to produce. Indeed, our quality assessments show that the rugs produced by treatment firms score significantly higher along virtually every quality dimension. At the same time, "unadjusted" productivity measures—those that do not control for rug specifications and quality (e.g., output per hour)—fall by 24–28% among treatment firms.

A simple theoretical framework shows that the quality upgrading we find is consistent with two distinct mechanisms that have not been disentangled in the literature to date. The first mechanism amounts to a movement along the PPF. We posit that the output per unit input of a firm depends on both rug specifications and an output efficiency parameter χ_a; high-specification rugs take longer to weave and, ceteris paribus, firms with higher χ_a produce more output per unit input. Quality also depends on rug specifications and a quality efficiency parameter, χ_q, and is increasing in both. The export opportunity exposes firms to buyers in high-income markets, and these buyers are willing to pay more for quality than domestic buyers. As long as firms find it profitable to do so, they will raise specifications, and hence improve quality. Under this first mechanism, firms already know how to manufacture high-quality rugs and the opportunity to export simply induces a movement along the PPF. That is, there is no change in either efficiency parameter.

A second mechanism involves an increase in the efficiency parameters induced by exporting: learning-by-exporting. This learning can come about through transfers of knowledge from buyers to producers, or from learning-by-doing if such learning would not have happened without exporting (a distinction we return to). Learning-by-exporting is an outward shift of the PPF which can occur either by raising χ_a (producing more output per input conditional on specifications) or raising χ_q (producing higher quality conditional on specifications). When these increases in efficiency are biased toward the production of high-quality rugs, both rug

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

quality and profits will rise. Of course, the two mechanisms are not mutually exclusive, but the presence of learning-by-exporting is important because it implies larger gains from trade.

We present five pieces of evidence to show that the improvements in performance come, at least in part, through learning-by-exporting.

The first is that both quality and productivity rise after adjusting for product specifications (recall that “unadjusted” productivity falls). If firms only moved along the PPF, specification-adjusted quality and productivity would remain constant. Second, to ensure specifications were fully controlled for, at the end of our experiment we asked all firms to manufacture an identical-specification rug for the domestic market using identical inputs and a common loom in a workshop that we leased (a “quality lab”). The rugs that treatment firms produced received higher scores along every quality metric and were more accurate in terms of the desired size and weight; moreover, treatment firms do not take longer to produce these rugs despite their higher levels of quality. Third, we explore the evolution of quality and productivity over time. Inconsistent with a movement along the PPF, where quality and productivity should immediately jump and then stay fixed, we document learning curves for both. Fourth, we draw on correspondences between foreign buyers and the intermediary, as well as on a log book of discussions between the intermediary and producers, to document that our results come, in part, from knowledge flows. In particular, we show that treatment firms improve quality most along the particular quality dimensions discussed during meetings between the intermediary and the producer, and that the vast majority of these discussions were about specific weaving techniques to improve quality. This, coupled with more anecdotal reports of knowledge passed from foreign buyers to the intermediary, suggests that the improvements in efficiency occur partly through knowledge transfers from intermediaries and foreign buyers. Fifth, we rule out investment explanations by showing that treatment firms make no monetary or time investments in upgrading; nor do they pay, even implicitly, for the knowledge they receive from the intermediary. Taken together, the evidence strongly supports the presence of learning-by-exporting.

As with any industry- or country-specific study, it is important to acknowledge issues relating to external validity. In terms of the context of our study, the firms in our sample are small—typically having only one full-time employee—and production is

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

not automated. Hence, our study has little to say about learning-by-exporting for large firms manufacturing complex products. Of course, it is precisely their small size that allows us to assemble a large sample necessary for inference; and the fact that they manufacture products using the same technology allows us to design more specific surveys and improves statistical power. The firms also export via an intermediary, rather than directly, but indirect exporting is common for the Egyptian rug industry and in other industries and countries.

In terms of the experiment itself, our treatment induces exporting by reducing matching frictions between firms and sophisticated foreign buyers. Such frictions are of interest both for theoretical and policy reasons. Allen (2014) estimates that matching frictions explain half of the overall variation in trade costs, while Lederman, Olarreaga, and Payton (2010) note that reducing matching frictions for small and medium-sized firms (SMEs) is a key goal for export promotion agencies. That said, our experiment does not reduce the trade frictions more typically studied in the trade literature such as tariffs or transport costs. While ultimately an empirical question requiring further research, we conjecture that exporting to high-income markets will generate similar learning for developing-country SMEs however they are induced to export (whether by reducing tariffs, trade costs, or matching frictions).

Two more caveats are necessary. Given the difficulties we document in generating orders and the implicit labor costs of our time, it is unclear if it is efficient for our sample firms to pay the fixed costs required to find sophisticated foreign buyers. The goal of this article is not to carry out a cost-benefit analysis of export facilitation programs or to isolate market failures preventing firms from exporting in the absence of assistance (both questions that would require an entirely different experimental design involving a large number of NGO-led interventions). Instead, the goal is to identify the presence of learning-by-exporting. Second, given the nature of our experiment, we are unable to distinguish exporting from selling to sophisticated domestic buyers—that is, domestic buyers who demand high quality and possess knowledge about how to achieve such quality. However, these buyers are scarce in developing countries, and as the literature on quality upgrading we discuss below suggests, the presence of such buyers may be the most pronounced difference between internal and external trade for developing-country firms. These two characteristics of

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

exporting—the fact that not only is knowledge transferred but there are also incentives in place to encourage firms to absorb and implement this knowledge—explain why exporting to high-income markets can be a particularly effective way of transferring technology to developing countries.

Our results relate to a number of papers that span the trade and development literatures. Most directly, we contribute to a voluminous literature that seeks to identify the existence of learning-by-exporting. The evidence from these studies is mixed,[2] in part due to the severe selection and measurement issues highlighted above. We directly confront selection through random assignment and directly confront measurement both by collecting very detailed data on the production process and by setting up a quality lab that allows us to perfectly control for product specifications. In doing so, we follow Syverson (2011) and Bloom and Van Reenen (2010) who advocate improving our understanding of productivity through more careful measurement.

Our findings also relate to the literature on quality upgrading. Studies using country- or product-level data show that product quality positively co-varies with export destination income per capita (Schott 2004; Hallak 2006, 2010); and firm-level studies suggest that exporting exposes firms in developing countries to sophisticated buyers who demand higher quality.[3] Unlike much of this literature that must infer quality from price data or certifications, or through structural models that back out quality from prices and quantities, we collect direct measures of quality.[4] In addition, the combination of our RCT, the quality lab and the rich

2. For example, Clerides, Lach, and Tybout (1998) and Bernard and Jensen (1999) conclude that firms self-select into export markets. In contrast, several papers using alternative approaches to deal with selection (e.g., matching estimators or instrumental variables) find some support for learning; see de Loecker (2007, 2013), Park et al. (2010), and Marin and Voigtlander (2013). Keller (2004), Wagner (2007), and Harrison and Rodriguez-Clare (2010) survey the literature.

3. For example, see Verhoogen (2008), Manova and Zhang (2012), Crozet, Head, and Mayer (2012), Brambilla, Lederman, and Porto (2012), Hallak and Sivadasan (2013), and Bastos, Silva, and Verhoogen (2014). In contrast, Marin and Voigtlander (2013) find that rather than quality rising, marginal costs decline because Colombian firms invest in reducing production costs as they enter export markets.

4. Papers that infer quality in these ways include Verhoogen (2008), Khandelwal (2010), Hallak and Schott (2011), Manova and Zhang (2012), and Feenstra and Romalis (2014).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

survey data allows us to contribute to this literature by showing quality upgrading occurs, at least in part, through improvements in technical efficiency rather than through movements along the PPF alone.

Finally, although the use of RCTs is novel in the trade literature, the methodology has been used to understand supply constraints in firms (e.g., de Mel, McKenzie, and Woodruff 2008, 2010, 2014; Bloom et al. 2013 explore credit constraints, input market frictions, and managerial constraints). We complement this literature by providing the first experimental evidence for the importance of demand constraints and the effects of relaxing those constraints through expanding market access.

The rest of the article is organized as follows. Section II describes the research setting. Section III explains our experimental intervention and introduces the data. Section IV examines the impact on profits and Section V decomposes the profit changes. Section VI presents a theoretical framework that then guides our five-step approach to detecting learning-by-exporting. Section VII concludes.

II. Research Setting

II.A. Finding a Viable Setting

To carry out a randomized evaluation of the impact of exporting, we partnered with Aid to Artisans (ATA), a U.S.-based NGO with a mission to create economic opportunities for producers of handmade products in developing countries. Discussions began in October 2009, and ATA had just received funding to implement a market-access program in Egypt.

ATA's program in Egypt followed their standard protocol for generating successful exporting relationships between small-scale developing-country producers and high-income markets. First, ATA explores the country in question for products that would both appeal to consumers in high-income markets and be priced competitively. Once candidate products are found, ATA identifies a lead intermediary based in the developing country.

The lead intermediary assists in finding small-scale producers to manufacture the products, is the conduit for passing information and orders between the producers and the buyers, and handles the export logistics required to ship the products to importers or retailers abroad. ATA then works closely with the

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

intermediary to design and market appealing products. Working through a lead intermediary firm, rather than matching individual producers directly with foreign buyers, is an important aspect of ATA's model. By matching local intermediaries with foreign buyers, the intermediaries can aggregate orders to spread the fixed costs of exporting across many small producers. The ultimate objective is to foster self-sustaining relationships whereby ATA can pull out and the local intermediaries maintain or even expand their export sales.

The process of exporting via an intermediary is common for handmade products and among small firms more generally. For example, Chinese customs data show that 52% of Chinese exports in the specific HS code that the rugs in our study are classified under—HS 570231 ("Carpets and other textile floor coverings, wool")—go through intermediaries (compared to 20% of overall Chinese exports).[5] Across sectors, World Bank Enterprise Surveys (2006–2014) reveal that 62% of exporters with five or fewer employees (and 36% of all exporters) use an intermediary to export.

Alongside ATA, we searched for viable Egyptian products and identified handmade carpets from Fowa as having potential. In terms of the industry choice, both the handmade craft industry and the rug industry are large and important sources of employment in Egypt, as well as in many other developing economies (see Online Appendix B). Turning to the location, Fowa is a periurban town with a population of 65,000 located two hours southeast of Alexandria. The town is well known for its carpet cluster, which contains hundreds of small firms that use wooden foot-treadle looms to manufacture flat-weave rugs—a product in which Egypt has a strong historical reputation. Crucially, we also identified a local firm, Hamis Carpets, to serve as the lead intermediary. Hamis is the largest intermediary in Fowa and accounts for around 20% of the market. At the time, Hamis earned 70% of its sales in the domestic market, mostly selling to distributors and retailers in Cairo, Alexandria, and Luxor.

The firms in Fowa typically consist of a single owner who operates out of a rented space or sometimes his (all producers in our sample are men) home. Family members or hired labor assist

5. These numbers come from the data described by Ahn, Khandelwal, and Wei (2011) who show that smaller firms use intermediaries to avoid large fixed costs associated with directly exporting.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

with setting up the loom and the finishing stage. The process of producing rugs is standardized across firms. The two key inputs beyond the loom and labor are warp thread—wool or cotton thread that spans the length of the rug and is not visible on the final rug but is necessary to hold the rug together; and weft thread—the colorful threads weaved between the warp threads using a shuttle. Online Appendix C provides additional details regarding the production process.

Firms self-identify as specialists in one of four flat-weave rug types: duble (the focus of this article), tups, kasaees, and goublan.[6] The average duble rug destined for domestic markets is sold by firms for LE42.5 (about US$7 at the prevailing exchange rate)[7] and requires 5.9 hours of labor, per m^2. After accounting for input costs, hourly wages are roughly LE3 ($0.48).

Within a particular rug type, quality can vary substantially. There are two determinants of quality. First, higher quality is associated with more demanding specifications. Specifications are codifiable attributes of the rug that are typically chosen by the buyer; for example, the number of colors, the thread count or the type of input thread. Figure I shows one such specification sheet from a foreign buyer in our experiment. Second, higher quality is associated with better weaving technique. For example, how flat the rug lies on a hard surface is determined by how skillfully the warp and weft threads are installed on the loom, and whether the threads are held correctly while weaving. Similarly, how well defined the corners are, how accurately the design was followed, and whether the rug adheres to the desired size specifications depend on weaving skill. These attributes of the rug are both difficult to codify and depend on the skill of the firm.

II.B. Generating Export Orders

It took the combination of ATA and Hamis more than two years to generate sustained export orders from clients in high-income countries. Generating sustained orders was not guaranteed. The textile market is competitive and conversations

6. Duble and tups rugs are the most common; kasaees rugs are the cheapest and woven from rags; goublan rugs are the most expensive and are works of art hung on walls. See Online Appendix Figure C.1 for examples.

7. As discussed later, there are two baseline surveys that were run in July 2011 and February 2013. The exchange rate on July 1, 2011, was LE5.94 and LE6.68 per US$1 on February 1, 2013; we apply the average value, 6.31, throughout.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

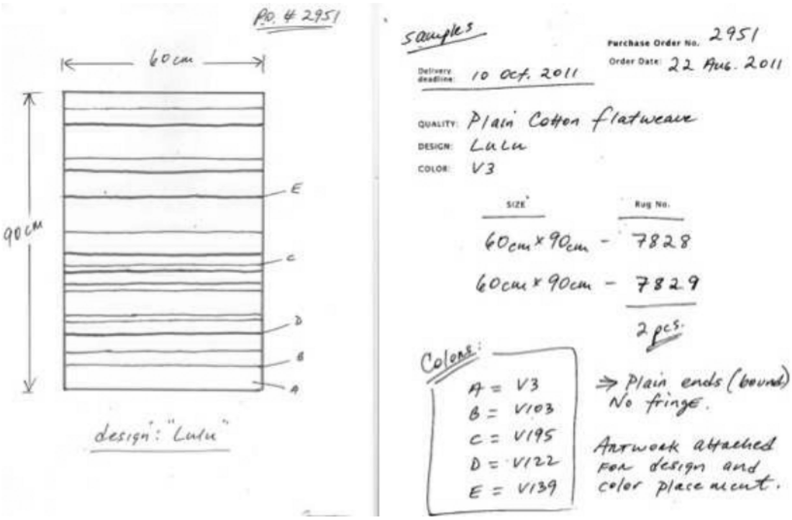

FIGURE I

Example of Rug Specifications Provided by a Potential Foreign Client

with ATA's staff revealed that only one in seven matches lead beyond trial orders. This is consistent with Eaton et al. (2013) who estimate that only one in five importer-exporter matches results in successful business relationships.

ATA first brought the CEO of Hamis to the United States for a training course that covered marketing and distribution. At the same time, ATA hired an Italian consultant to design rug samples. Hamis was then provided with marketing support—both by displaying these samples at various international gift fairs and by directly introducing Hamis to foreign importers or retailers through a U.S.-based rug intermediary.

Hamis and potential foreign buyers would discuss pricing, delivery time, and product specifications (design, colors, materials, and so forth; recall Figure I provides an example of how these specifications are codified after these discussions). Hamis would then organize the production of sample orders, either from its in-house weavers or from one of the treatment firms in our sample.

The majority of rugs demanded by foreign buyers are duble rugs, although one client ordered kasaees rugs. There have been no orders for goublan rugs, even though the local market in Egypt perceives these rugs to require the most skilled weaving techniques; the painting-like style of goublan rugs is unlikely to

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

FIGURE II

Examples of Domestic Rugs and Export Rugs

Figure provides examples of the domestic rug (left) and export rugs ordered by foreign buyers (center, right). The domestic rug is the duble rug that firms were asked to manufacture at the quality lab.

appeal to buyers in high-income countries (see Online Appendix Figure C.1 for an example). Likewise, domestic-style duble rugs did not attract interest from abroad. Instead, it appears that foreign buyers prefer "modern" designs, as illustrated in Figure II.

After one and a half years of searching, in June 2012 Hamis Carpets secured its first large export order (3,640 m^2) from a German buyer. As of June 2014, multiple buyers continue to place large, regular orders. Figure III reports that cumulative export production between December 2010 and June 2014 (the end of our experiment) totaled 33,227 m^2, resulting in cumulative payments to the producers of LE982,351 ($155,682).[8] As described in the next section, these orders were entirely sourced from our treatment firms, which forms the basis of our experiment.

III. The Experiment

III.A. Experimental Design

We designed the following export-market-access intervention. We drew a sample of small rug producers (described in more

8. The revenues received by the intermediary from these orders were significantly larger. As shown in Online Appendix Table F.2 and explained further below, Hamis reports applying a 33% markup on export orders after paying for materials.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

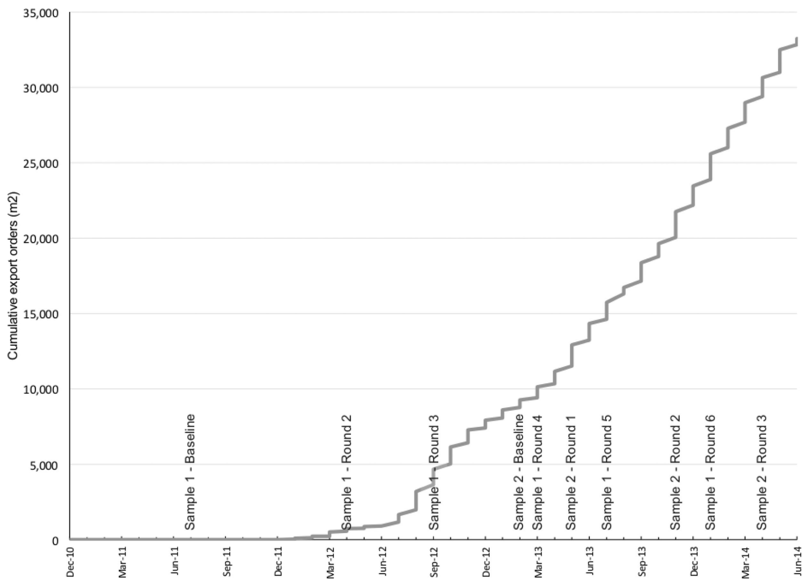

FIGURE III

Cumulative Export Orders

detail in the next section). The firms were divided into two groups, treatment and control.

As described, the local intermediary, Hamis Carpets, secured export orders with ATA's marketing assistance. Treatment firms were then visited by a representative of Hamis Carpets and provided the opportunity to fill an initial export order. More precisely, Hamis Carpets showed them the rug design, explained that the carpet would be exported to a high-income country, and offered them an order of 110 m^2, which translates to about 11 weeks of work. The 110 m^2 was chosen by trading off the desire to have a reasonably sized initial order and the need to have enough orders to treat the firms. Hamis was free to choose the price offered to the producers based on the specifications of the rugs (we analyze prices later). To ensure all rug orders were consistent across producers, Hamis provided the input thread and two loom components that are design-specific, the reed and the heddle (Online Appendix C describes these components).[9] At the same time, as is typical in buyer-producer relationships, Hamis would discuss the

9. It is standard practice in Fowa for intermediaries to provide design-specific reeds and heddles, and the cost to the intermediary does not depend on the design.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

technical aspects of the specific rug order and answer any questions the firm may have. Firms would deliver rugs to Hamis with payment on delivery.

As further export orders were generated, we tried to mimic a normal buyer–seller relationship as closely as possible. Hamis was allowed to continue to place orders with treatment firms but was not bound to make subsequent purchases from firms whose quality was below par or who could not deliver on time. (Firms were informed of this arrangement.) In other words, the experiment protocol simply required Hamis to offer an initial order to the treatment firms. In contrast, the control firms were not contacted by Hamis Carpets about an initial order, and Hamis was forbidden from sourcing from them through the duration of the experiment.[10] The experimental protocol is described in more detail in Online Appendix D.1.

As we show later, very few firms had ever knowingly exported at baseline. Thus, the intervention provided treatment firms with the opportunity to produce rugs for the export market. Comparing outcomes between treatment and control firms allows us to identify the causal effects of being provided with such an opportunity.[11]

An alternative experiment would be to provide our control firms with a similar quantity of rug orders but from domestic rather than foreign sources. We did not pursue this approach for reasons both theoretical and practical. Trade models typically model exporting as a demand shock (e.g., Krugman 1979; Melitz 2003), sometimes with features distinct from domestic demand shocks. Increasing demand is also the primary motivation for many export promotion policies (e.g., sending trade delegations or analyzing opportunities for domestic firms in foreign markets). Therefore, to assess the impacts of exporting, it is natural to

10. A project coordinator and Fowa-based survey team ensured that the protocols were followed. However, one control firm was incorrectly treated due to an error by Hamis. In the empirical analysis we make the most conservative assumption and keep this firm in the control group.

11. Although the initial order is random, subsequent orders are not. Hence, treatment effects are potentially heterogeneous if some firms are better able to take advantage of the initial opportunity than others. If there was a constraint on the total export orders, better firms may receive subsequent orders at the expense of worse firms. Under the reasonable assumption that learning is concave in export orders (an assumption supported by the learning curves we estimate later), this will lead us to underestimate the impacts of the experiment compared to a setting with unconstrained orders. To mitigate this concern, we restricted the size of the treatment group in Sample 2 to ensure that constraints on offering orders to all treatment firms would not be binding.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

include the increased demand it brings. In terms of the practical limitations, if we were to provide equally-sized initial domestic orders, it is unclear on what dimension they should be equal given the different profit margins and hours required per rug. And then it would have been almost impossible to match subsequent foreign and domestic orders over time, not least because of the fluctuations in the Egyptian economy that would have made it extremely difficult to obtain $155,682 of firm orders from new domestic sources.

III.B. Sample Details and Takeup

To obtain the sample of firms, we carried out a recruitment drive in Fowa in July 2011. To be eligible, the firm had to have fewer than five employees, work on their own account (meaning that they bought their own inputs when an order required), and have never previously worked with Hamis. The recruitment drive generated a sample of 303 firms that specialized in one of the four rug types described in Section II.A.

Anticipating that we would not secure orders for every type of rug, we stratified the sample on the type of rug produced and the loom size (which determines the maximum width of rug that can be produced). Within each stratum we randomized firms into treatment and control.

For reasons that will become clear momentarily, we refer to these 303 firms as Sample 1. Table I shows the total number of firms by rug type and treatment status.

It proved difficult to secure sufficient export orders to treat every firm in the treatment group with an 110 m^2 initial order. As detailed in Section II.B, we were only successful in generating large and sustained orders for one of the four rug types, duble rugs, and even then only one and a half years after the baseline survey. For this reason, we were only able to offer firms the opportunity to produce duble rugs, and we could only offer multiple smaller orders of 10–20 m^2 spread out over many months instead of the full 110 m^2 order in one go. This resulted in very low takeup (16%) among specialized producers of the nonduble rug types (firms not taking up cited an unwillingness to switch rug types; see Online Appendix D). Even among the 79 specialist duble producers, only 14 of the 39 treated firms firms took up, with nontakeup firms citing an unwillingness to jeopardize their existing dealer relationships for the small 10–20 m^2 order sizes.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE I
SAMPLE AND TAKEUP STATISTICS

	Duble orders					Kasaees orders
	Joint sample	Sample 1			Sample 2	Sample 1
Statistic	Duble firms (1)	Goublan firms (2)	Tups firms (3)	Duble firms (4)	Duble firms (5)	Kasaees firms (6)
Firms	219	103	83	79	140	38
Treatment firms	74	49	42	39	35	19
Takeup firms	46	5	8	14	32	5
Initial packet size (m^2)	110	110	110	110	110	250
Successful takeup firms	46	4	6	14	32	5
Mean output conditional on takeup (m^2)	538	586	589	778	434	303
Std. dev. of output conditional on takeup	188	174	208	132	177	76

Notes. Table reports statistics by firm type and sample. The first row displays the number of firms within each rug type and sample. The second row displays the number of firms in the treatment group. The third row indicates the number of firms who accepted the treatment and agreed to make rugs for export. The fourth row is the initial order size (in square meters) offered to each takeup firm. The fifth row shows the number of firms that completed the initial order successfully and received subsequent orders from Hamis. The sixth row indicates average output conditional on takeup. The last row reports the standard deviation of output in the first year, conditional on takeup.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

Once large and sustained export orders for duble rugs arrived in June 2012, it was feasible to offer the full 110 m^2 initial order in one go. Given the opportunity to implement the experimental design we originally intended, we asked our surveyors to locate every remaining duble firm in town that satisfied our sample criteria. We found 140 additional duble firms in this manner which we refer to as Sample 2. We stratified on loom size as before, and randomly selected 35 firms for the treatment group (with the number of treated firms determined by Hamis's capacity constraints). Consistent with the claim that takeup in Sample 1 was low due to the small order sizes, column (5) of Table I shows that 32 of the 35 treated Sample 2 firms took up when offered the full initial order in one go. All of the duble firms in both Sample 1 and 2 that took up the opportunity were "successful" in the sense that they both delivered the 110 m^2 that constituted the initial order and then received subsequent orders from Hamis.

Given that we only secured large and sustained export orders for duble rugs and that very few nonduble firms were willing to manufacture this rug type, we were essentially unable to treat the nonduble strata and so exclude them from our analysis. To be clear, if the focus of this article were to simply evaluate an export-market-access program, it would be important to understand why the intervention only generated sustained exports for one of the four products. Instead, we use the random variation our experiment generates to investigate a long-standing question in international trade: does exporting improve firm productivity?

For the analysis in the rest of the article, we combine the firms in the duble strata from Sample 1 with the firms in Sample 2—who are exclusively duble producers—to form the joint sample of 219 firms (74 in treatment, 47 of which took up). This is essentially all the duble producers in Fowa willing to identify themselves to our surveyors (minus 10 firms that refused to participate). For completeness, Online Appendix G presents the results for duble firms in the two samples separately,[12] as well as results for the nondouble strata.

12. It is important to note that although the two samples look very similar in terms of characteristics (see Online Appendix Table F.1), Sample 1 received a fundamentally different treatment—a treatment of multiple smaller export orders spread over time—than did Sample 2. Nevertheless, other than the lower takeup, results are similar across the two samples. Out of 43 comparisons of treatment-on-the-treated coefficients between the two samples, three are significantly different at the 10% level and one is significant at the 5% level (in line with the number of Type 1 errors we would expect).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE II
SURVEY TIMELINE

Survey timeline	Sample 1	Sample 2
Baseline round 0	July–Aug 2011[a]	Feb–Mar 2013[a]
Follow-up round 1	Nov–Dec 2011[b]	May–June 2013
Follow-up round 2	April–May 2012	Nov–Dec 2013
Follow-up round 3	Sept–Dec 2012	May–June 2014[a]
Follow-up round 4	Mar–Apr 2013	
Follow-up round 5	July–Oct 2013[a]	
Follow-up round 6	Jan–Mar 2014	
Quality lab	June 2014	June 2014

Notes. Table reports the timeline for the data collection on duble firms by sample.
[a]Supplementary questions about household and firm outcomes are included in both samples' baseline round 0 surveys, in follow-up round 5 for Sample 1, and follow-up round 3 for Sample 2.
[b]The Egyptian survey company we hired for the Sample 1 baseline trained a new enumerator for the first Sample 1 follow-up. Unfortunately, we discovered that this enumerator did not actually interview all of the firms and entered in fake data for some, so we discard the data from follow-up round 1 for Sample 1. We immediately fired the enumerator and hired new employees in January 2012 who conducted all subsequent surveys. We managed these employees directly and implemented a stricter auditing procedure, as well as back checks using external evaluators, to ensure data integrity.

III.C. Data

We collected multiple rounds of data at around four-month intervals. In total, Sample 1 was interviewed seven times and Sample 2 four times.

The baseline round occurred before treatment firms were provided with the opportunity to export and included questions on (a) firm production, (b) rug quality, and (c) household and demographic characteristics. We repeated this full survey for follow-up round 5 of Sample 1 and follow-up round 3 of Sample 2. In all other follow-up rounds we administered an abbreviated survey on firm production and rug quality. The survey timeline is shown in Table II.

The production module records production activity for the month preceding the survey interview. All nominal variables are converted to real values using the official Egyptian CPI. We collect measures of profits, revenues, expenses, output quantity and prices, input quantity and prices, total labor hours worked, and the specifications of the rugs produced that month. These specifications include: (a) the type of rug being produced, (b) how difficult the rug is to make rated on a 1–5 scale by a master artisan accompanying the surveyors (see below), (c) the amount of weft

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

thread used per m^2 of the rug (thread count), (d) the type of weft thread used (e.g., Egyptian wool, cotton), (e) the number of colors used in the rug, and (f) which segment of the market the rug is aimed at as reported by the master artisan (normal, mid, or high).

The quality module records the quality of the rugs being produced by firms at the time of the survey. Rug quality is assessed by a master artisan under our employ who is a well-known and respected member of the rug community in Fowa. Quality was measured along 11 dimensions: (a) corners, (b) waviness, (c) weight, (d) touch, (e) packedness, (f) warp thread tightness, (g) firmness, (h) design accuracy, (i) warp thread packedness, (j) inputs, and (k) loom.[13] The quality along each dimension is scored on a 1 to 5 scale, with higher numbers denoting higher quality. These quality metrics capture differences across rugs that are vertical in nature; for example, at equal prices, both foreign and domestic consumers would prefer a flatter-lying rug or a more accurate design. As discussed in Section II.A, higher quality scores reflect a combination of better specifications and greater weaving skill.

For takeup firms, a second quality module is available at higher frequency. These firms deliver rugs to Hamis on a weekly basis. On receiving each batch of rugs, Hamis records the size accuracy, design accuracy, packedness, and weight accuracy.

We collected a third set of quality and productivity measures in June 2014 by asking firms to manufacture an identical domestic rug using identical inputs and a common loom in a work space we had rented (our quality lab, see Section VI.C for further details). Production was timed and on completion the rugs were measured, anonymized, and then sent to be scored along the same

13. The Sample 1 baseline survey recorded six quality metrics to which we subsequently added five more metrics. Corners captures the straightness of where the rug edges meet. Waviness captures how flat the rug lies when placed on a hard surface. Weight captures how close the actual weight of the rug is to the intended weight. Touch reflects the feel of the rug. Packedness measures how well the rug holds together (poorly packed rugs have small holes). Warp thread tightness measures the tightness of the warp thread which helps determine how tightly held the weft thread is. Firmness measures the firmness of the rug when held. Design accuracy captures how accurate the design is to the intended pattern. Warp thread packedness measures how visible the warp thread is (it should not be visible at all). Inputs measures the quality of the input threads. Loom measures the quality of the loom.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

quality dimensions by both the master artisan and a professor of Handicraft Science from Domietta University.

Finally, we collect information on knowledge flows—the number of visits Hamis made to firms and what was discussed during each visit—that we discuss in Section VI.E.

III.D. Summary Statistics

Table III shows baseline balance between the treatment and control groups. The table reports regressions of each variable on a treatment dummy and strata fixed effects, and reports the constant (the mean of the control firms) and treatment coefficient (the difference between control and treatment means). Panel A shows summary statistics for the household characteristics of the firm owner. The mean age in the control group is around 51 years and, on average, firm owners have slightly more than 37 years of experience working in the rug industry. Roughly 63% of firm owners are illiterate. The average household size is 4.2 and the average total monthly household income from all activities is LE1,090 ($173).

Panel B reports statistics from the rug business. Monthly profits from the rug business average LE646 ($102) in the control group. Firms report 247 labor hours in the previous month, which amounts to around 22 days of work at 11 hours a day. As noted earlier, firm sizes are small because this was an explicit criterion in choosing our sample: the average firm has just over one worker. Total output per month is 50 m^2 and only about 12% of firms have ever knowingly produced rugs for the export market. The final row of Panel B reports the average rug quality across the 11 dimensions.

Across both panels we find no statistical differences between treatment and control firms with one exception: treatment firms report lower quality scores at baseline. The final row of Table III reports attrition across survey rounds (Online Appendix Table F.3 reports attrition by round). Attrition has been relatively low with a non-response rate of approximately 11% per round, which does not vary across treatment and control groups, while attrition in the quality lab is similar at 14% and balanced across groups.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE III
BASELINE BALANCE

	Control group mean	Difference in treatment	N
Panel A: Household characteristics			
Age	51.0	0.9	218
	(0.7)	(1.6)	
Number of years in rug business	37.7	0.2	213
	(0.8)	(1.7)	
Illiterate?	0.63	0.10	214
	(0.03)	(0.07)	
Household size	4.2	0.0	219
	(0.1)	(0.2)	
Household income	1090.0	76.5	219
	(91.2)	(228.0)	
Digit span recall	5.8	0.2	204
	(0.1)	(0.2)	
Panel B: Firm characteristics			
Price per square meter	30.2	6.8	218
	(3.3)	(7.8)	
Direct monthly profits from rug business	646	7.9	218
	(41.8)	(81.5)	
Reported monthly profits from rug business	806	−10.4	217
	(38.5)	(84.4)	
Hours worked last month	247	−1.7	218
	(5.6)	(11.7)	
Number of employees	1.09	0.0	218
	(0.0)	(0.1)	
Total produced last month (m^2)	50.0	3.3	218
	(4.3)	(10.0)	
Ever exported?	0.12	0.02	219
	(0.02)	(0.05)	
Average quality	2.63	−0.13***	218
	(0.03)	(0.05)	
Joint F-test	1.23		
Attrition in follow-up surveys	0.11	0.00	815
	(0.01)	(0.02)	
Attrition in quality lab	0.14	0.02	219
	(0.03)	(0.05)	

Notes. Table explores baseline balance. Each row is a regression of the named variable on a constant, treatment dummy and strata fixed effects; the constant (control group mean) and treatment dummy are reported. The third-to-last row reports the F-test for a regression of the treatment dummy on all 14 baseline balance variables. Profits and prices are winsorized at the 1st and 99th percentile to trim outliers (without winsorizing, the sample still remains statistically balanced between treatment and control groups). The final rows report average attrition rates across all survey rounds and in the quality lab respectively. Significance: *.10; **.05; ***.01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

IV. Causal Impacts of Export-Market Access on Profits

IV.A. Empirical Specifications

The randomization methodology allows us to use a straightforward specification to assess the impact of the export-market access on firm profits:

$$y_{it} = \alpha_1 + \beta_1 Treatment_i + \gamma_1 y_{i0} + \delta_s + \tau_t + \varepsilon_{it}, \tag{1}$$

where y_{it} is the profit measure, $Treatment_i$ is an indicator variable that takes the value 1 if firm i is in the treatment group, τ_t are time period fixed effects, δ_s are strata fixed effects and y_{i0} is the value of the dependent variable at baseline.[14] We are essentially combining all follow-up survey rounds to increase precision, and we cluster standard errors at the firm level to take account of the fact that errors may be correlated within firms. As equation (1) controls for the baseline value of the dependent variable, we cannot include observations from the baseline survey in the regression.[15] Since not all firms who were offered the opportunity to export took up that offer, equation (1) is an intent-to-treat (ITT) specification.

We also present results from the treatment-on-the-treated specification (TOT) which scales up the treatment effect to take

14. We note that any effects we find will be attenuated if there were spillovers to control firms. However, we find no support for geographic spillovers between treatment and control firms (see Online Appendix Table F.4).

15. Alternatively we could use all survey rounds, include firm fixed effects, and interact $Treatment_i$ with a postbaseline dummy. We prefer our specification since if the dependent variable is measured with noise and not strongly autocorrelated, as is the case for business profits (an autocorrelation of 0.33 among our control firms), the fixed effects estimator will perform more poorly in an experimental study than the analysis of covariance estimator in equation (1); see McKenzie (2012). For comparison we also report the key results using firm fixed effects in Online Appendix Table F.5 and find that results are very similar. When the baseline value is missing for a firm, our specification drops that firm, leading the number of observations to vary across estimates. In cases where the baseline value is missing for an entire strata, we include the firm but code the missing value as 0. We find no evidence that missing values at baseline are correlated with treatment when testing the eight main outcomes of the article (i.e., those in Online Appendix Table F.5).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE IV
IMPACT OF INTERVENTION ON FIRMS KNOWINGLY EXPORTING

	(1) ITT	(2) TOT
Indicator for ever exported	0.55***	0.76***
	(0.06)	(0.07)
R-squared	0.33	0.45
Control mean	0.20	0.20
Observations	191	191

Notes. Table regresses an indicator for if a firm has ever knowingly produced rugs for export markets on indicators for treatment (column (1)) or takeup (column (2)). The question was asked in round 5 for Sample 1 and round 3 for Sample 2. The TOT regression instruments takeup with treatment. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Significance: *.10; **.05; ***.01.

account of the fact that not everyone took up the treatment:[16]

$$y_{it} = \alpha_2 + \beta_2 Takeup_{it} + \gamma_2 y_{i0} + \delta_s + \tau_t + \nu_{it}, \qquad (2)$$

where $Takeup_{it}$ takes the value 1 if a firm took up the opportunity to export. This is a time-varying measure that turns on when a firm first produces carpets for the intermediary and stays on subsequently. Of course takeup is not random and may be correlated with unobservables, and so we instrument $Takeup_{it}$ with the variable $Treatment_i$ that is uncorrelated with the error (and the baseline control) thanks to the randomization procedure.

Before showing results on profits and other metrics, we first show that indeed the intervention worked, insofar as treatment firms were more likely to manufacture rugs for export markets. To do so, we replace y_{it} with a dummy variable that takes the value 1 if a firm ever knowingly made rugs for export. As shown in Table IV, being in treatment raises the probability of ever exporting by 55 percentage points from a baseline of 13%. We also report the TOT specification, which suggests even more dramatic increases.[17]

16. The TOT will be an upper bound if the firms who took up the intervention were the ones with most to gain from exporting (although we find no evidence of selection into takeup based on observables).

17. Note that the ITT and TOT do not perfectly scale up by the takeup rates shown in Table I since a handful of firms that eventually took up had not yet done so by the first follow-up rounds.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

IV.B. Profit Results

Following de Mel, McKenzie, and Woodruff (2009)—who assess the performance of a variety of methods to elicit profits from small firms—we construct four profit measures (described in more detail in Online Appendix E). The first measure directly asks firm owners to report profits from the previous month (excluding wage payments to themselves). This is the measure most strongly advocated by de Mel, McKenzie, and Woodruff (2009) because it avoids error due to the timing mismatch between revenues and expenses. The second measure constructs profits from two survey questions that ask firms to report their total revenues and total costs from the previous month. The third measure constructs profits from the production modules that contain detailed information on prices and quantities of inputs and outputs. The fourth measure, hypothetical profit, is based on a question that asks firms how much profit they would earn from selling rugs if they purchased a specific quantity of inputs.

Table V, Panel A shows the results of running specifications (1) and (2) on logged values of the four profit metrics, and for each we report the ITT and TOT.

Columns (1) and (2) report the specifications using the (log) direct monthly profit measure. The ITT coefficient is 0.26 and significant at the 1% level, implying that the export treatment increases monthly profits by approximately 26%. The TOT coefficient is, not surprisingly, larger at 42% and also statistically significant. Columns (3) and (4) report specifications using the profit measure constructed from total revenues and costs in the previous month and columns (5) and (6) report specifications using the profit measure constructed from prices and quantities. The ITTs are 21% and 19%, respectively, very similar to the direct profits results. This similarity suggests that any timing mismatch between revenues and expenses is not severe in our setting where firms store little inventory, and that rug inputs are not diverted to household consumption. Finally, we examine the hypothetical profit measure in columns (7) and (8). Although these estimates are slightly higher (the ITT is 37%), it is reassuring that all four measures increase by economically and statistically significant amounts. We also note that the treatment effects reflect profits rising more among treatment firms rather than profits falling among control firms; control firm profits increased in real terms across baseline and post-baseline survey rounds (regressing log

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE V
IMPACT OF EXPORTING ON FIRM PROFITS

	Log direct profits		Log (reported revenues – reported costs)		Log (constructed revenues – constructed costs)		Log hypothetical profits	
	(1) ITT	(2) TOT	(3) ITT	(4) TOT	(5) ITT	(6) TOT	(7) ITT	(8) TOT
Panel A: Profits (in month prior to survey)								
Treatment	0.26***	0.42***	0.21***	0.37***	0.19***	0.34***	0.37***	0.68***
	(0.05)	(0.08)	(0.06)	(0.10)	(0.06)	(0.10)	(0.11)	(0.19)
R-squared	0.21	0.22	0.16	0.18	0.16	0.18	0.19	0.19
Control mean (in levels)	929	929	931	931	951	951	541	541
Observations	573	573	644	644	685	685	687	687
Panel B: Profits per owner hour (in month prior to survey)								
Treatment	0.20***	0.32***	0.17***	0.29***	0.16***	0.28***	0.25***	0.46***
	(0.05)	(0.08)	(0.05)	(0.09)	(0.05)	(0.09)	(0.07)	(0.12)
R-squared	0.14	0.14	0.12	0.13	0.13	0.13	0.19	0.18
Control mean (in levels)	3.53	3.53	3.54	3.54	3.55	3.55	5.56	5.56
Observations	573	573	637	637	684	684	687	687

Notes. Table reports treatment effects on different measures of real profits in the month prior to the date of the survey, all measured in logs. See text for descriptions of each measure. Dependent variable in Panel A is profits. Dependent variable in Panel B is profits per owner hour. Owner hours include the hours of family member production when recorded. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Control group means are reported in levels in Egyptian pounds (LE) in Panel A and LE/hour in Panel B. The TOT regressions instrument takeup with treatment. Standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

profits of control firms on a postbaseline dummy yields a coefficient of 0.470; std. err. of 0.190).

These regressions indicate that the export treatment causally increases measured profits by between 19% and 37%. Of course, profits may have risen partly because firms increased their labor hours. This is an issue for our profits measures since most firms are owner-operated and the profit measures do not account for the implicit wages paid to the owner. If we focus on profits per owner hour, these concerns are mitigated since the value of the owners' time will be similar across treatment and control due to the randomization.[18] Therefore, Table V, Panel B examines profits per owner hour by dividing each profit variable by the total hours worked by the owner (or other unpaid family members when recorded) in the previous month. Using the direct profit per owner hour measure in columns (1) and (2), we find that the ITT estimate is 20% and again significant at the 1% level. This estimate is lower than the corresponding estimates for profits, which implies that owners of treatment firms worked more hours. The remaining columns also show lower estimates. However, the basic message remains the same: the opportunity to export raised profits per owner hour by 16–25%. The differences between the panels suggest that total owner hours increased by around 5%, a result we confirm in the next section when we analyze labor hours.

Before turning to mechanisms, and in particular whether these improvements in firm performance occur through learning-by-exporting, we note that it is not surprising that providing firms with a demand shock increases profits. What is surprising is the magnitude of the effect. Many supply-side interventions on similar samples of firms have had limited profit impacts. A recent literature, surveyed by McKenzie and Woodruff (2013), has carried out impact evaluations of business training programs for small firms. These programs had a statistically significant impact on profits in only two out of nine studies that measured profits. Another popular supply-side intervention is expanding access to credit.

18. Of course this measure will still be misleading if owners of treatment firms increased their effort per hour worked. We were unable to devise survey questions that could accurately capture worker effort, but we did ask about job stress measured on a scale of 1 to 5. Reassuringly, when we regress job stress on log output per hour, treatment and the interaction of the two (plus round and strata fixed effects), the coefficients are neither individually nor jointly significant (F-stat of 1.40).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

The literature on the impacts of credit on profits for small firms also finds mixed results. For instance, de Mel, McKenzie, and Woodruff (2008) find returns to capital of around 5% a month while Banerjee (2013) cites several credit interventions that produced no statistically significant increases in profits. One possible interpretation of these mixed results and our findings of substantial increases in profits is that supply-side interventions may only be effective where there are no constraints on demand. Thus, demand-side interventions such as our market-access program may be complementary to more-standard supply-side interventions.

V. Sources of Profit Changes

V.A. *Prices, Output, and Inputs*

This section explores the proximate sources of the increase in profits. To fix ideas consider the following profit function for a firm:

$$\max_{l,k} \pi = px(l,k) - wl - rk - F, \tag{3}$$

where p is the price a firm receives for one unit of rug. The quantity of rugs produced is x, w is the wage paid for each hour of labor l, r is the rental rate on capital k, and F is a fixed cost of production. Although we analyze inputs below, we do not include input costs in equation (3) since a large majority of firms (91%) receive raw material inputs from their intermediary and hence do not pay for these expenses.[19]

Table VI uses our survey data to examine these various components of profits. Columns (1) and (2) evaluate the impact of the intervention on the log output price. The ITT specification indicates a 43% increase in prices with the opportunity to export while the TOT indicates a 78% increase. Thus, part of the profit increase from exporting is coming from significantly higher prices per m^2. Columns (3) and (4) examine the impact of the opportunity to export on the log total output weaved by the firm in the previous month (measured in m^2 and unadjusted for product specifications). The ITT estimate is −26% while the TOT is −47%; there

19. For the subset of firms that do purchase inputs themselves, we subtract the prices of the warp and weft thread inputs from p to make these prices comparable across all firms.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE VI

SOURCES OF CHANGES TO FIRM PROFITS: COMPONENTS OF PROFITS

	Log output price $\left(\frac{LE}{m^2}\right)$		Log output (m^2)		Log hours worked		Number of employees		Log number of looms		Log warp thread ball (kg)	
	(1) ITT	(2) TOT	(3) ITT	(4) TOT	(5) ITT	(6) TOT	(7) ITT	(8) TOT	(9) ITT	(10) TOT	(11) ITT	(12) TOT
Treatment	0.43***	0.78***	−0.26***	−0.47***	0.05**	0.08**	0.01	0.01	−0.02	−0.04	0.15***	0.25***
	(0.10)	(0.19)	(0.09)	(0.17)	(0.02)	(0.04)	(0.01)	(0.01)	(0.04)	(0.06)	(0.05)	(0.08)
R-squared	0.16	0.15	0.24	0.22	0.12	0.13	0.02	0.02	0.13	0.13	0.24	0.24
Control mean (in levels)	28.2	28.2	64.1	64.1	269.0	269.0	1.0	1.0	1.1	1.1	6.0	6.0
Observations	691	691	676	676	678	678	695	695	694	694	600	600

Notes. Table reports treatment effects on output prices and quantities, hours, number of employees (inclusive of owner), looms, and the size of the warp thread ball (which is a proxy for the length of the production run), all measured in logs except number of employees. The TOT regressions instrument takeup with treatment. Hours worked are calculated using average daily hours and number of days worked last month. Control group means are reported in levels. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Standard errors are clustered by firm. Significance: * .10; ** .05; *** .01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

is a large decline in output in treatment firms relative to control firms.

Columns (5)–(10) document the impact of the intervention on firm scale, as captured by labor and capital usage. Columns (5) and (6) show the log of total hours l worked by all employees in the firm in the previous month. The ITT estimate indicates a labor increase of 5% and the TOT estimate is 8%. This increase in labor hours comes on the intensive margin: as shown in columns (7) and (8) there is no change in the number of employees (inclusive of the owner). Since most firm owners are the primary weavers, and helpers are often family members, we have very few observations of the wage w that may also be responding to the opportunity to export. (We already showed that profits per owner hour increase but this combines the shadow wage with firm profits.) In contrast to labor, we find no increase in capital usage k, as measured by the log number of active looms in use (columns (9) and (10)). The fact that expansion occurs primarily along the intensive margin suggests there may be large nonconvexities associated with hiring additional workers, particularly since an additional weaver is likely to need his own loom and requires owners to manage a full-time employee for the first time.

Finally, we turn to fixed costs F in columns (11) and (12). We try to capture economies of scale via a proxy for the length of the production run. Firms place a warp thread ball on the loom at the beginning of a production run. A larger warp thread ball enables firms to amortize the costs of restringing the loom over more units. We find that the size of the warp thread ball increases by 15% in the treatment group, indicating that the opportunity to export lowers fixed costs through longer production runs that require less frequent restringing.

Table VII examines input prices and quantities. As noted already, most firms do not purchase the material inputs, but we did ask these firms to estimate the price of the weft and warp thread inputs. The first two columns use these data to explore the impact of the intervention on thread prices. Reported weft thread prices increase 20%. In contrast, there is no evidence that warp thread prices are higher among treatment firms. These two findings are sensible given the production technology. The warp thread is critical to maintain the rug structure but is not observable in the finished rug. Meanwhile, the weft thread is observable and can vary by both material type (cotton, wool, polyester, etc.), material grade (e.g., Egyptian wool or New Zealand wool) and

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE VII

SOURCES OF CHANGES TO FIRM PROFITS: INPUTS

	Log weft thread price $\left(\frac{\text{LE}}{\text{m}^2}\right)$		Log warp thread price $\left(\frac{\text{LE}}{\text{m}^2}\right)$		Log weft thread quantity (g)		Log warp thread quantity (g)	
	(1) ITT	(2) TOT	(3) ITT	(4) TOT	(5) ITT	(6) TOT	(7) ITT	(8) TOT
Treatment	0.20***	0.33***	−0.04	−0.07	−0.19**	−0.34**	0.04	0.06
	(0.06)	(0.10)	(0.03)	(0.06)	(0.10)	(0.17)	(0.11)	(0.20)
R-squared	0.22	0.24	0.27	0.27	0.23	0.22	0.10	0.11
Control mean (in levels)	12.8	12.8	18.1	18.1	110.0	110.0	17.8	17.8
Observations	564	564	685	685	677	677	686	686

Notes. Table analyzes input prices and quantities, all measured in logs. The TOT regressions instrument takeup with treatment. Control group means are reported in levels. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Standard errors are clustered by firm. Significance: * .10; ** .05; *** .01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

thickness. Note that although columns (5)–(8) suggest that input quantities (measured in grams) do not increase with the opportunity to export, the output decline implies that the rugs produced by treatment firms use more material inputs and are heavier than those produced by control firms.

The increases in prices, labor input usage, and the length of production runs appear consistent with two workhorse models used to study international trade. Comparative advantage models, such as the Ricardian model, would predict that export prices are higher for products that Egypt has a comparative advantage in (and it is reasonable to think handmade flat-weave rugs are such a product). In this framework, the opportunity to export would also raise the quantity of labor being used in rug production, as we find. Similarly, our findings on scale and fixed costs are consistent with a standard scale effects story whereby exporting enables firms to reach larger markets and spread fixed costs over more units (e.g., Krugman 1979). However, the reduction in output is not consistent with either of these frameworks. The results are also not consistent with exporting simply being a generic demand shock (which would also yield an increase in output).

The reductions in output accompanied by rising output (and input) prices point to export-induced quality upgrading. If high-quality rugs require more labor inputs, rug output can fall alongside increasing revenues and input usage. If high-quality rugs also require more expensive inputs (as shown by Kugler and Verhoogen 2012), the rise in material input prices provides further evidence of quality-upgrading. In the next subsection we confirm this conjecture.

V.B. Quality and Unadjusted Productivity Measures

We first draw on the detailed quality metrics described in Section III.C to confirm that treatment firms are indeed manufacturing higher quality products. We have 11 different quality metrics that are ranked on a 1–5 scale with 5 being the best for that dimension of quality.

Table VIII presents the quality results. Instead of implementing specifications (1) and (2) separately for each quality metric, we regress a stack of all 11 metrics on interactions of the treatment (or takeup for the TOT) with indicator variables for each quality metric. We also include interactions of the quality metric indicators with baseline values, strata fixed effects and round fixed

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE VIII
IMPACT OF EXPORTING ON QUALITY LEVELS

	Control mean	(1) ITT	(2) TOT
Panel A: Quality metrics			
Corners	2.98	1.11***	1.70***
		(0.12)	(0.11)
Waviness	2.99	1.10***	1.68***
		(0.12)	(0.10)
Weight	3.08	1.07***	1.63***
		(0.11)	(0.11)
Touch	3.12	0.40***	0.66***
		(0.06)	(0.07)
Packedness	3.11	0.89***	1.59***
		(0.11)	(0.12)
Warp thread tightness	3.05	0.83***	1.49***
		(0.10)	(0.12)
Firmness	2.98	0.87***	1.60***
		(0.11)	(0.12)
Design accuracy	3.17	0.79***	1.41***
		(0.10)	(0.12)
Warp thread packedness	3.05	1.07***	1.65***
		(0.11)	(0.11)
Inputs	3.07	0.89***	1.62***
		(0.10)	(0.12)
Loom	2.02	0.03	0.05
		(0.02)	(0.04)
R-squared		0.44	0.60
Observations		6,885	6,885
Panel B: Stacked quality metrics			
Stacked quality metrics	2.96	0.79***	1.35***
		(0.09)	(0.08)
R-squared		0.39	0.54
Observations		6,885	6,885

Notes. Panel A stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality-metric indicator variable. The coefficients on the interactions provide the treatment effects separately for each metric. The TOT instruments takeup interacted with quality metric with treatment interacted with quality metric. Each regression includes baseline values of the quality metric, strata and round fixed effects, and each of these controls interacted with quality-metric. Panel B constrains the treatment effects to be equal across quality metrics; these regressions include baseline values, strata and round fixed effects. Control group means are reported in levels. Standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

effects. The resulting coefficients are identical to those from running separate regressions for each quality metric, but run this way we can cluster standard errors by firm to account for any firm-level correlations within quality metrics across time or across quality metrics within a period.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

For 10 of the 11 quality metrics, quality is significantly higher among treatment firms (all at the 1% level). The one exception is loom quality. The lack of a treatment effect on loom quality is consistent with our understanding of the technology for rug production. Although the loom size determines the maximum rug width, it matters little for rug quality.

Since it is difficult to parse all 11 quality metrics separately, Panel B of Table VIII restricts the coefficients on the treatment dummy to be identical across all 11 quality metrics (recall they were all run in a single stacked regression).[20] Given the previous results, it is not surprising that we obtain positive and statistically significant ITT and TOT estimates when we do this. On average, quality (on a scale of 1 to 5) is 0.79 point higher among treatment firms. These are substantial increases in quality given a standard deviation of quality of 0.55 at baseline.

We also examine two simple productivity measures that do not adjust for changes in product specifications or quality: unadjusted output per labor hour and unadjusted total factor productivity (TFP).

Unadjusted output per labor hour comes from firms' responses to the question: "how long does it take you to make 1 m^2?" The second measure also accounts for capital inputs (although recall there is limited variation in capital across firms with 92% of firms using only one loom).[21] Specifically, unadjusted TFP is equal to the residual from an estimated Cobb-Douglas production function that includes both labor and capital (see Appendix for a full description of the procedure).

Table IX shows that both of these productivity measures fall in treatment relative to control. Looking at the ITTs, unadjusted output per hour is 24% lower and the unadjusted TFP is 28% lower, with even larger TOT effects.

V.C. Quality-Upgrading Mechanisms

The finding that quality rises and unadjusted productivity falls alongside rising profits is consistent with two different

20. This method is similar to estimating the impact of treatment on a standardized index of quality (e.g., Kling, Liebman, and Katz 2007), but we prefer our method as it produces more conservative estimates in our data (i.e., higher standard errors).

21. Looms do vary by size but we control for loom sizes through strata fixed effects in the analysis below.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE IX
IMPACT OF EXPORTING ON UNADJUSTED PRODUCTIVITY

	Log unadjusted output per hour		Log unadjusted TFP	
	(1) ITT	(2) TOT	(3) ITT	(4) TOT
Treatment	−0.24***	−0.42***	−0.28***	−0.50***
	(0.09)	(0.16)	(0.09)	(0.16)
R-squared	0.18	0.16	0.26	0.24
Control mean (in levels)	0.26	0.26	0.49	0.49
Observations	687	687	674	674

Notes. Table reports treatment effects for the two productivity measures: log unadjusted output per labor hour (in $\frac{m^2}{hour}$) and log unadjusted TFP. See text and Appendix for the methodology used to obtain unadjusted TFP. The TOT specifications instrument takeup with treatment. Control group means are reported in levels. Regressions control for baseline values of the variable, round and strata fixed effects. Standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

quality-upgrading mechanisms, and the distinction is important for understanding how exporting improves firm performance. In the first mechanism, firms always knew how to manufacture the high-quality rugs demanded by rich-country buyers. If foreign buyers pay higher prices, but particularly so for high-quality products, firms will upgrade quality as long as the returns offset any costs (e.g., more expensive inputs or more labor inputs).

This is a movement along the PPF. Under this mechanism, the export opportunity raises the relative price of high-quality rugs and profit-maximizing firms respond by producing rugs with specifications associated with high quality. What does not change through this mechanism is technical efficiency.

Although it is challenging to provide a direct mapping between profit margins and quality levels, we provide some suggestive evidence for this phenomenon by analyzing Hamis's (self-reported) cost structure for domestic and foreign orders. Hamis reports 9% profit margins on (lower-quality) domestic orders and substantially higher margins of 33% on (higher-quality) foreign orders. (The full cost structure is broken down in Online Appendix Table F.2). Although these are Hamis's profit margins, not the firms', if there is profit sharing there would be a similar relationship between profit margins and rug quality for producers.

A second mechanism is learning-by-exporting, which we follow the literature and define as an export-induced change in

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

technical efficiency (Clerides, Lach, and Tybout 1998; de Loecker 2007). This is a shift out in the PPF and can include both transfers of information from buyers to producers and learning-by-doing that would not have happened in the absence of exporting (e.g., if export products have steeper learning curves). If such changes in technical efficiency are biased toward the production of high-quality rugs, quality upgrading can also occur through these learning processes.

We emphasize that these two mechanisms are not mutually exclusive. In fact, a rise in the price of quality is potentially a precondition for the learning-by-exporting described above. In these contexts, where the opportunity to export raises the price of quality, learning-by-exporting generates further profit increases beyond those from simply moving along the PPF. In the next section, we define learning-by-exporting more precisely and provide evidence that it is present in our setting.

VI. Detecting Learning-by-Exporting

VI.A. *A Framework for Detecting Learning-by-Exporting*

To be explicit about the learning-by-exporting mechanism, we enrich the profit function as follows:

$$\max_{l,k,\lambda} \pi = px - wl - F \tag{4}$$

$$x = a(\lambda; \chi_a) f(l, k) \tag{5}$$

$$q = q(\lambda; \chi_q) \tag{6}$$

$$p = p_0 + bq(\lambda). \tag{7}$$

Rug output x and rug quality q are determined by separate production functions. Prices, p, are determined by a price function that is exogenous to the firm and is increasing in the quality of the rug, with $b > 0$ determining the price of quality.

Each of the two production functions depends on efficiency parameters, χ_a and χ_q, that capture the skill of the firm, as well as on a choice variable: the product specifications of the rug indexed by λ. (Recall that specifications are codifiable rug attributes that buyers agree on before ordering; see Figure I for an example of

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

such an agreement.) High-λ rugs have more demanding specifications, in the sense that they require more inputs, primarily labor hours, to produce, and we assume that these high-λ specifications are also associated with high-quality rugs.

More precisely, the production function for output x has two components. Labor and capital inputs are mapped to output through $f(l, k)$ and output per unit input is determined by the function $a(\lambda; \chi_a)$, a TFP metric that is "unadjusted" for rug specifications.[22] Unadjusted TFP $a(\cdot)$ is decreasing in λ, since rugs with more demanding specifications require more inputs, and increasing in the efficiency parameter χ_a, which governs how quickly a firm produces rugs of a particular set of specifications with a given set of inputs. Collecting these two derivatives:

$$\frac{\partial a(\lambda; \chi_a)}{\partial \lambda} < 0 \qquad \frac{\partial a(\lambda; \chi_a)}{\partial \chi_a} > 0. \tag{8}$$

Quality is determined by the function $q(\lambda; \chi_q)$ which we assume is increasing in product specifications as quality is achieved in part through more demanding specifications. In addition, quality increases in the efficiency parameter χ_q which governs a firm's ability to make quality given a particular set of specifications. Collecting these two derivatives:

$$\frac{\partial q(\lambda; \chi_q)}{\partial \lambda} > 0 \qquad \frac{\partial q(\lambda; \chi_q)}{\partial \chi_q} > 0. \tag{9}$$

With this structure in hand, it is straightforward to clarify what constitutes a movement along the PPF due to exporting and what constitutes a shift out (i.e., learning-by-exporting). Firms move along the PPF when there is an increase in b, the price of quality (due to, for example, foreign buyers' higher willingness to pay for quality). This leads firms to choose higher specifications λ, and by expression (9), quality rises. In contrast, learning-by-exporting occurs when exporting raises χ_a and/or χ_q, the two efficiency parameters, and hence shifts out the PPF. As mentioned earlier, this process can occur as firms move into high-quality products with steep learning curves or through transfers of knowledge from foreign buyers to domestic producers. We expect transfers of knowledge about quality, χ_q, to be particularly relevant for firms in

22. We abstract from material inputs in the production function since, as discussed earlier, intermediaries typically provide the raw materials to firms.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

low-income countries that sell to buyers in high-income countries since these buyers are likely to demand such high-quality products and possess knowledge about how to produce them. Despite the different theoretical implications, we are unaware of earlier work that seeks to distinguish these two quality-upgrading mechanisms.

To see that this theoretical framework can generate reductions in unadjusted TFP alongside improvements in quality through either mechanism, we rearrange the total derivatives of the first order conditions with respect to λ and l. First, consider an increase in the price of quality b. As long as there are diminishing marginal returns to raising specifications (i.e., concavity of q and a in λ), firms choose to raise specifications and so equilibrium quality q^* rises and equilibrium unadjusted TFP a^* falls:[23]

$$\frac{dq^*}{db} = q_\lambda \frac{d\lambda}{db} > 0 \qquad \frac{da^*}{db} = a_\lambda \frac{d\lambda}{db} < 0.$$

Now consider an increase in the quality efficiency parameter χ_q. As long as the complementarity between λ and χ_q in producing quality is sufficiently large—that is, weaving skill is particularly valuable for the production of high-specification rugs, $\frac{\partial^2 q}{\partial\lambda\partial\chi_q} > 0$ (e.g., higher thread-counts require more dexterity to weave)—firms also choose to raise specifications and so q^* rises and a^* falls:[24]

$$\frac{dq^*}{d\chi_q} = q_{\chi_q} + q_\lambda \frac{d\lambda}{\partial\chi_q} > 0 \qquad \frac{da^*}{d\chi_q} = a_\lambda \frac{d\lambda}{d\chi_q} < 0.$$

Intuitively, the strong complementarity ensures that it is profitable for the firm to raise specifications since quality, and hence price per m^2, increases faster than a declines. The same pattern can arise with an increase in the output efficiency parameter, χ_a, as long as there is a sufficiently large complementarity between λ and χ_a in producing output—$\frac{\partial^2 a}{\partial\lambda\partial\chi_a} > 0$; for example, skill matters more for speed when producing less-familiar high-specification

23. More precisely, $\frac{d\lambda}{db} = \frac{-(qa)_\lambda}{b(qa)_{\lambda\lambda} + p_0 a_{\lambda\lambda}} > 0$ if $q_{\lambda\lambda} < 0$ and $a_{\lambda\lambda} < 0$, where subscripts denote partial derivatives.

24. Specifically, we require that qa be supermodular in (λ, χ_q) for $\frac{d\lambda}{d\chi_q} = \frac{-b(qa)_{\lambda\chi_q}}{b(qa)_{\lambda\lambda} + p_0 a_{\lambda\lambda}} > 0$ (in addition to the concavity in λ).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

rugs—to ensure that the slowdown from raising specifications exceeds the direct increase in speed from higher χ_a.[25]

Our model does not allow for investments that raise the χ parameters. For the production function for physical output, x, investments that increase output should be fully captured by increases in some type of capital or labor in a well-specified production function. If the returns to such investments rise with the opportunity to export, any resulting changes should not be classified as learning-by-exporting under our definition since these potential investments were already accounted for in the PPF. (A similar argument can be made for the quality production function if we amend equation (6) to include labor and capital.) Hence, purchasing a more efficient weaving machine or paying for a training course in response to the export opportunity would not be considered learning-by-exporting. In contrast, tacit knowledge passed on by a buyer or intermediary which is neither anticipated nor paid for by the firm, even implicitly, would be.[26] Such a categorization is consistent with the learning-by-exporting literature that considers these types of knowledge transfers archetypal.

Empirically detecting learning-by-exporting is challenging for two reasons. First, firms with high efficiency parameters are likely to self-select into export markets making it difficult to disentangle treatment effects of exporting from selection (Melitz 2003).

The most convincing analyses to date rely on matching techniques, which require that researchers fully specify the underlying selection model (e.g., see de Loecker 2007). Here, we exploit the randomization to ensure that the opportunity to export is uncorrelated with initial levels of χ_a and χ_q.

Second, even if self-selection were not an issue, researchers typically measure technical efficiency through residual-based TFP. TFP measures that do not adjust for prices (which is rarely the case) may suggest learning-by-exporting when firms are just moving along the PPF or obtain a higher markup in export

25. Specifically, we require $(pa)_{\lambda\chi_a} > \frac{a_{\chi_a}}{a_\lambda}(pa)_{\lambda\lambda}$ for both $\frac{dq^*}{d\chi_a} = q_\lambda \frac{d\lambda}{d\chi_a} > 0$ and $\frac{da^*}{d\chi_a} = a_{\chi_a} + a_\lambda \frac{d\lambda}{d\chi_a} < 0$ to be satisfied (in addition to the concavity in λ), where $\frac{d\lambda}{d\chi_a} = \frac{-b(qa)_{\lambda\chi_a} - p_0 a_{\lambda\chi_a}}{b(qa)_{\lambda\lambda} + p_0 a_{\lambda\lambda}}$.

26. For this reason our framework excludes marketing capital as a factor in the production function. We believe, consistent with the learning-by-exporting literature, that the knowledge generated from matching with foreign buyers is not anticipated by the firms and hence should not be reflected in productivity. See Appendix for further discussion.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

markets. In the few cases where price adjustments are made, measuring quantity-based TFP requires comparing products with identical specifications and quality levels. This is typically achieved by focusing on homogeneous goods like concrete and block ice (e.g., Foster, Haltiwanger, and Syverson 2008) where trade frictions and learning may be limited; or by categorizing products based on administrative classifications and using a demand model to infer quality from prices and market shares (e.g., de Loecker et al. 2016). In contrast, we exploit our rich panel data and quality lab to solve these measurement issues.

We test several implications of the framework to detect learning-by-exporting:

- Step 1 We use our detailed data on product specifications to show that although our unadjusted productivity measure—corresponding to $a(\cdot)$ above—falls with the opportunity to export (recall Table IX), specification-adjusted productivity rises—consistent with χ_a increasing. We also show that specification-adjusted quality rises, consistent with χ_q increasing. If there is no learning-by-exporting, specification-adjusted productivity and quality should be unchanged as there is no change in the efficiency parameters.
- Step 2 We demonstrate that when asked to produce identical-specification domestic rugs using the same loom and the same inputs, treatment firms produce higher quality products and do not take longer to do so. Again, if there is no learning-by-exporting, treatment and control firms should produce identical-specification domestic rugs at the same quality.
- Step 3 We use time-series data to establish that quality and productivity evolve over time as cumulative export production increases, consistent with a learning process. In contrast, if firms simply moved along the PPF we would expect a discontinuous jump upon exporting as firms immediately move to new quality and productivity levels.
- Step 4 We draw on correspondences between foreign buyers and Hamis, as well as a log book of discussions between Hamis and the firms, to document that our results come, in part, from knowledge transfers (information that would be irrelevant if firms were only moving along the PPF). In particular, we show that treatment firms improve

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

quality most along the particular quality dimensions that are discussed during meetings with Hamis. This evidence also strongly suggests that learning-by-exporting is not solely driven by learning-by-doing (triggered by the export orders), but in part through transfers of knowledge.

Step 5 We rule out alternative hypotheses that there were adjustment costs, scale effects, or that firms made investments to raise output and quality that were not adequately captured by our labor and capital measures. In particular, we show that treatment firms make no monetary or time investments in upgrading, and do not pay, even implicitly, for the knowledge they receive from the intermediary.

VI.B. Step 1: Conditioning on Rug Specifications

If firms are only moving along the PPF, changes to unadjusted productivity and quality should occur only through changes in rug specifications: $\frac{da}{db}|_\lambda, \frac{dq}{db}|_\lambda = 0$. That is, producers always knew how to produce the particular rugs demanded by foreign buyers, but previously chose not to because domestic buyers did not value these rugs. If there is learning-by-exporting, then we would expect productivity and/or quality to rise, conditional on rug specifications, due to an increase in χ_a or χ_q: $\frac{da}{d\chi_a}|_\lambda, \frac{dq}{d\chi_q}|_\lambda > 0$.

To separate these two hypotheses, we repeat the quality and productivity regressions above but control in various ways for the specifications of the rug being manufactured at the time of the survey visit. Recall from Section III.C that we have six dimensions of rug specifications. Although imperfect, we note that many studies simply control for product differences through product fixed effects based on statistical classifications.

Our first specification—the type of rug—is the analogous control, although it uses a much finer classification than standard trade classifications (e.g., all of our seven rug types would fall within the U.S. HS 10-digit classification 5702311000). The remaining specifications, such as thread count or design difficulty, are rarely observed by researchers. Such controls are possible because there is overlap in rug specifications across firms selling to domestic and foreign markets. This overlap can be seen in Figure IV, which plots the distribution of each of the six specifications separately for firms that are producing rugs for export (i.e., $Takeup_{it} = 1$) and those that are not.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

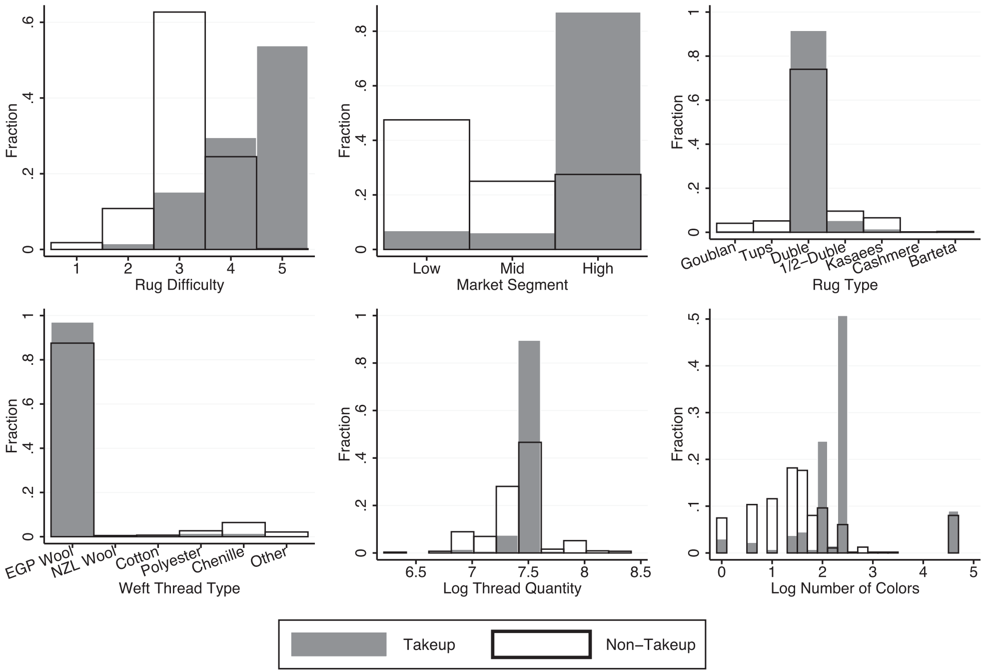

FIGURE IV

Overlap in Rug Specifications between Domestic and Export Orders

Figure plots the density of the six rug specifications for takeup (shaded) and nontakeup (outline) firms.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

Note that if our specification controls are very crude, that will tend to bias our findings toward the unconditional results we found in Tables VIII and IX. Hence, the prediction that productivity should rise conditional on specifications is particularly informative since unadjusted productivity falls. (Step 2 deals more directly with the possibility that the specification controls are imperfect.)

We present three sets of results. The first approach regresses quality and the two unadjusted productivity measures on treatment (or treatment instrumented with takeup) as in Section V.B but now also includes controls for the six specifications and their baseline values.

The second approach goes further toward ensuring that the treatment and control firms we compare are making identical rugs by including fixed effects for each of the 435 unique combinations of the six specifications. Since approximately one third of firm-round pairs are making unique rugs, the cost of this approach is that we lose a significant number of observations.

The third approach follows from equation (5) which suggests that we can directly infer χ_a from the residual of a production function estimation that includes specifications. As discussed in more detail in the Appendix, we use control firms to estimate this production function and then calculate specification-adjusted TFP for each firm and round using the estimated coefficients on labor, capital, and specifications. We construct similar measures for quality and output per hour by regressing these variables on rug specifications in the control group and use the resulting coefficients to construct adjusted metrics (actual minus predicted).

This procedure also partially addresses a second issue. Although treatment is exogenous by design, the specification controls in the first two approaches may be endogenous. Of course, if higher-ability firms selected into higher-specification rugs in the control group, the coefficients on rug specifications will be biased. However, if anything, this bias will lead us to find no productivity gain.[27]

27. Specifically, due to this selection, the productivity penalty for making a high-specification rug will be larger than the coefficients imply. If our experiment induced lower-ability firms to make high-specification rugs, the ITT that compares specification-adjusted productivity between treatment and control would likely be biased downward (because lower-ability firms in treatment would "appear" less productive when using the biased coefficients for the adjustment).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

Table X, Panel A reports the approach with specification controls, Panel B the approach with fixed effects for specification combinations, and Panel C the approach with specification-adjusted dependent variables. Before discussing the effects on quality and productivity, it is reassuring to note that the specification controls in Panel A have the signs we assumed in the model: more difficult rugs are associated with higher quality and lower unadjusted productivity, while those destined for lower segments of the market are associated with lower quality and higher unadjusted productivity. The R-squared rises substantially compared to the regressions without specification controls in Tables VIII and IX (increasing from 0.39 to 0.64 in the quality ITT, and from 0.18 and 0.26 to 0.57 and 0.62 in the productivity ITTs); this suggests that the rug specifications have substantial explanatory power.

Turning to the treatment effects in Table X, recall that without conditioning on specifications (i.e., Panel B of Table VIII and Table IX) quality rises and the productivity measures fall. Conditioning on specifications using the three approaches described above, quality again rises significantly but the signs on the productivity measures flip from significantly negative to significantly positive in all three cases. (The productivity ITTs correspond to productivity increases between 14% and 31%.) That is, conditional on making similar rugs, treatment firms are making them faster than control firms. These results suggest a rise in the efficiency parameters χ_a and χ_q.

VI.C. Step 2: Production of Identical-Specification Domestic Rugs (the Quality Lab)

The second step exploits our experimental setting to compare quality and productivity across firms producing identical-specification domestic rugs (rather than relying on specification data to control for the type of rug). If firms are only moving along the PPF, when asked to make rugs with identical specifications, quality and productivity should not differ across treatment and control firms (since treatment was randomly assigned). To carry out this test we brought the owners of each firm to a rented workshop in June 2014 and asked them to produce an identical-specification rug using identical inputs and the same loom. We chose rug specifications that mimicked a popular rug design sold at mid-tier domestic retail outlets in Egypt (the rug is shown

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE X
QUALITY AND PRODUCTIVITY CONDITIONAL ON SPECIFICATIONS (STEP 1)

	Stacked quality metrics		Log output per hour		Log TFP	
	(1) ITT	(2) TOT	(3) ITT	(4) TOT	(5) ITT	(6) TOT
Panel A: Specification controls						
Treatment	0.32***	0.78***	0.18**	0.44**	0.14**	0.35**
	(0.04)	(0.08)	(0.08)	(0.18)	(0.07)	(0.16)
(log) Thread quantity	0.04	0.02	−0.12	−0.13	−0.07	−0.08
	(0.05)	(0.04)	(0.13)	(0.13)	(0.13)	(0.12)
Difficulty control	0.47***	0.34***	−0.14***	−0.21***	−0.16***	−0.22***
	(0.02)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)
(log) # colors	0.03**	0.01	−0.05*	−0.07**	−0.06**	−0.07***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)
Low-market segment	−0.19***	−0.08**	0.43***	0.49***	0.42***	0.47***
	(0.03)	(0.03)	(0.08)	(0.09)	(0.07)	(0.08)
Mid-market segment	−0.19***	−0.05	0.29***	0.36***	0.26***	0.32***
	(0.04)	(0.04)	(0.08)	(0.09)	(0.07)	(0.08)
Rug type FEs	Yes	Yes	Yes	Yes	Yes	Yes
Input thread type FEs	Yes	Yes	Yes	Yes	Yes	Yes
R-squared	0.64	0.67	0.57	0.58	0.62	0.63
Observations	6,820	6,820	673	673	660	660
Panel B: Specification fixed effects						
Treatment	0.13**	0.53***	0.31***	1.26***	0.25***	0.94***
	(0.05)	0.17	(0.08)	(0.45)	(0.08)	(0.36)
Specification FEs	Yes	Yes	Yes	yes	Yes	Yes
R-squared	0.71	0.72	0.69	0.62	0.70	0.68
Observations	6,820	6,820	428	428	416	416
Panel C: Specification-adjusted dependent variables						
Treatment	0.42***	0.72***	0.18**	0.33***	0.20***	0.36***
	(0.05)	(0.04)	(0.07)	(0.13)	(0.07)	(0.12)
R-squared	0.18	0.27	0.06	0.10	0.13	0.18
Observations	6,860	6,860	678	678	669	669

Notes. Table reports treatment effects for the stacked quality measures and the two productivity measures after including various controls for the specifications of the rug on the loom at the time of the survey. The TOT specifications instrument takeup with treatment. In Panel A, there are both rug and input thread type fixed effects in addition to the specification controls included in the table. Panel B uses fixed effects for each of the 435 combinations of the six specification controls. Both panels use unadjusted productivity measures. Panel C uses specification-adjusted measures for quality and productivity; see text and Appendix for details. Regressions in all panels also control for baseline values of the dependent variable (and baseline values of the specification controls in Panel A), as well as round and strata fixed effects. Standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

in Figure II) and told the firms the orders were for a new buyer in Cairo. The master artisan assigned a difficulty rating of 3 for this rug (below the 4.28 average rating of export orders). We hired a new staff member to implement the quality lab and gave identical instructions to treatment and control firms. Owners were given a fixed payment a little above the market wage to compensate for producing in an external location. We provide the experimental protocol used for this step in Online Appendix D.3.

As discussed in Section III.C, each completed rug was given an anonymous identification number and the master artisan was asked to score it along 9 of the 11 quality dimensions discussed previously (we provided the inputs and the loom so the last two dimensions were not relevant). The identification system ensured that the master artisan had no way of knowing whether the rug was made by a treatment or control firm. We also sent the rugs to be scored by a second external quality assessor, a professor of handicraft science at Domietta University located two hours from Fowa, to cross check the accuracy of the master artisan's scoring.

Table XI, Panel A reports ITT and TOT results separately for each of these nine quality metrics.[28] Using the master artisan's scores, quality is significantly higher among treatment firms for all dimensions. Reassuringly, treatment firms also score significantly higher along every dimension using the professor's quality assessments.

Panel B constrains the coefficients on treatment to be the same across all the quality metrics. Unsurprisingly, this average effect is positive and statistically significant. In terms of magnitudes, the point estimate from the master artisan's scores is 0.64 and the standard deviation is 0.75 among the control group, implying that the opportunity to export increases quality levels by 0.85 standard deviations. For the professor's scores, the increase is 0.5 standard deviations.

Panel C reports the accuracy of rugs in terms of the length, width, and weight that we requested. We define these variables as the negative of the absolute deviation from the target value, so higher values reflect greater accuracy. Treatment firms produce rugs that are closer to the requested length and weight. We do not observe statistical differences in the width of the rugs, but this is

28. As before, we account for correlations across quality metrics by stacking our metrics, interacting treatment with each metric and a strata fixed effect, and clustering standard errors by firm.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE XI

QUALITY AND PRODUCTIVITY ON IDENTICAL-SPECIFICATION DOMESTIC RUGS (STEP 2)

	Master artisan			Professor		
	Control mean	(1) ITT	(2) TOT	Control mean	(3) ITT	(4) TOT
Panel A: Quality metrics						
Corners	3.23	0.72***	1.05***	3.31	0.29**	0.43**
		(0.14)	(0.17)		(0.13)	(0.18)
Waviness	3.17	0.55***	0.80***	3.31	0.25**	0.36**
		(0.14)	(0.18)		(0.12)	(0.16)
Weight	3.60	0.62***	0.91***	3.64	0.58***	0.86***
		(0.13)	(0.16)		(0.17)	(0.25)
Packedness	3.30	0.77***	1.14***	3.28	0.28**	0.42***
		(0.13)	(0.15)		(0.11)	(0.15)
Touch	3.29	0.52***	0.76***	3.27	0.36***	0.52***
		(0.11)	(0.14)		(0.12)	(0.16)
Warp thread tightness	3.00	0.51***	0.74***	3.30	0.25**	0.36**
		(0.09)	(0.11)		(0.12)	(0.16)
Firmness	3.21	0.71***	1.04***	3.23	0.29**	0.43***
		(0.14)	(0.17)		(0.12)	(0.16)
Design accuracy	3.65	0.53***	0.77***	3.45	0.27**	0.40***
		(0.11)	(0.15)		(0.11)	(0.15)
Warp thread packedness	3.05	0.87***	1.28***	3.20	0.39***	0.58***
		(0.14)	(0.17)		(0.12)	(0.16)
R-squared		0.21	0.34		0.11	0.14
Observations		1,680	1,680		1,667	1,667
Panel B: Stacked quality metrics						
Stacked quality metric	3.28	0.64***	0.94***	3.33	0.33***	0.48***
		(0.10)	(0.12)		(0.10)	(0.13)
R-squared		0.19	0.32		0.09	0.13
Observations		1,680	1,680		1,667	1,667

Panel C: Objective metrics	Control mean	(1) ITT	(2) TOT
Length accuracy	−4.51	1.43***	2.09***
		(0.51)	(0.71)
Width accuracy	−2.29	0.17	0.25
		(0.29)	(0.41)
Weight accuracy	−221.0	89.1***	131.0***
		(20.3)	(29.6)

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE XI
(CONTINUED)

	Control mean	(1) ITT	(2) TOT
Time (in minutes)	247.0	−5.67	−8.3
		(6.6)	(9.5)
R-squared		0.84	0.84
Observations		748	748

Notes. Table reports ITT and TOT specifications using the nine quality metrics from the quality lab. Panel A stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality-metric indicator variable. The coefficients on the interactions provide the treatment effects separately for each metric. The TOT instruments takeup (interacted with quality metric) with treatment (also interacted with quality metric). Panel B constrains the treatment effects to be equal across quality metrics. Columns (1) and (2) report scores from the master artisan. Columns (3) and (4) report scores from the professor of Handicraft Science at Domietta University. Panel C reports objective accuracy measures, which are calculated as the negative of the absolute error for that specification, so that a higher value indicates that the manufactured rug was closer to intended length (140 cm), width (70 cm), and weight (1,750 g). It also includes the time spent to produce the rug in minutes. As in Panel A, these are run in a single regression by intereacting the objective measure with treatment or takeup. All regressions include interactions of strata fixed effects with quality-metric indicators, and standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

expected since the loom size determines the width (and all firms used the same loom).

Finally, we recorded the time taken to produce the rug. Since the rug specifications, material inputs, and loom are identical for all firms in this setup, the time taken reflects firm productivity. The fourth row of Panel C shows that, on average, firms took four hours to produce the rug. Although the ITT is not significant, treatment firms took six minutes less. That is, despite manufacturing rugs with higher quality metrics, treatment firms spend if anything less time weaving, not more.

In the absence of learning-by-exporting, we would not expect differences between treatment and control firms when producing identical-specification rugs for the domestic market using the same inputs, the same loom, and at the same scale. If anything we might expect control firms to produce these rugs more quickly or at higher quality since they have recent experience manufacturing domestic designs and specifications. It also seems unlikely that treatment firms put more effort into weaving the rug because they were worried poor performance would jeopardize their relationship with Hamis. Firms were not informed of any link between the quality lab and Hamis Carpets—recall we hired a new staff member to run the lab to disassociate it from the export opportunity randomization as much as possible and we told firms that the order was from a new buyer in Cairo—and if firms did believe

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

there was a link it is just as plausible that control firms put in extra effort to impress Hamis to gain export orders.

In contrast, we find strong evidence of higher quality levels among treatment firms that persist even when manufacturing rugs for the domestic market, indicative of an increase in χ_q. As treatment firms do not take longer to produce these rugs, these results imply that a broader productivity measure that adjusts for both specifications and quality would rise substantially. Note that the lack of a significant treatment effect for time taken does not contradict the increase in productivity we found in Step 1. In the presence of a complementarity between λ and χ_a, the additional skill acquired through exporting will only translate into faster production for more demanding export rugs and not for these simpler domestic rugs that they were already familiar with.

VI.D. Step 3: Learning Curves

The third step examines the time paths of quality upgrading. Unlike a movement along the PPF, which should be instantaneous (see Section VI.F for a discussion of adjustment costs), learning processes typically take time. In the most obvious formulation, the efficiency parameters change with the opportunity to export through the cumulative production of export rugs. This captures the idea that efficiency improves with repeated interactions with buyers and/or because learning curves are steeper among export rugs that are less familiar to our sample firms. Therefore, if there is learning-by-exporting, productivity and quality are likely to rise with cumulative exports.

If there is no learning, although quality may immediately jump with the first export order (or unadjusted productivity may fall), the levels should remain constant with subsequent export orders.

To investigate potential learning curves in a nonparametric manner, we carry out a two-stage procedure. In the first stage, we regress our quality or productivity measures on both firm and round fixed effects.[29] In the second stage, we plot a kernel-weighted local polynomial regression of the residuals against cumulative export production. Since cumulative export production is only available for takeup firms, we only include these firms in

29. We use firm fixed effects here rather than baseline controls so that we can visualize the changes between baseline and the follow-up survey rounds which would not be possible with baseline controls.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

the second stage (although all firms are included in the first stage when we demean by survey round). Online Appendix F presents similar plots using the partially linear panel data estimator proposed by Baltagi and Li (2002).

Figure V shows these residual plots for the productivity measures as well as the stacked quality measure.[30] The upper and middle left plots show the two unadjusted productivity measures; the plots indicate a decline in productivity until about 600 m^2 after which it starts to rise. The second column uses the specification-adjusted measures described in Section VI.B. Both specification-adjusted output per hour and TFP rise with cumulative exports, consistent with the initial dip in unadjusted productivity being driven by the move to more difficult product specifications demanded by foreign buyers. The bottom row of Figure V presents the analogous learning curves for the stacked quality measures.

Both plots show a rise in quality up until 200 m^2 of exports and then a leveling off. The typical firm weaves 10–15 m^2 per week, which suggests that firms learn how to produce the quality demanded by foreigners within about five months. Overall, the learning curves suggest faster learning about quality efficiency χ_q than about output efficiency χ_a.[31]

In contrast to the ITT results in Steps 1 and 2, the fact we allowed Hamis to allocate follow-up orders means that more orders may have been given to firms whose quality or productivity was improving. With that caveat in mind, we see the evolution of productivity and quality among takeup firms as suggestive

30. Analyzing the Joint Sample is complicated here by the different timelines across samples. Online Appendix Figure F.4 shows the cdf of total export production across the two samples and Online Appendix Figure F.5 plots days since first order against export production for each firm. As Sample 1 firms started exporting earlier they drive all the variation at high values of cumulative exports but not at lower values, leading to a spurious discontinuity around this transition. Accordingly, Figure V restricts attention to the common support; the range of cumulative exports achieved by Sample 2 firms ($0-727$ m^2). Online Appendix Figures F.1 and F.2 present the two samples separately across their full ranges. See Online Appendix Table F.6 for a parametric analysis that interacts treatment with round dummies and finds most learning occurring by the first follow-up round.

31. The finding that learning about quality occurs quickly is consistent with other recent studies. In a randomized study of management practices in Indian textile firms, Bloom et al. (2013) find reductions in quality defects after just 10 weeks. Likewise, Levitt, List, and Syverson (2013) document a 70% decline in defect rates in an automobile manufacturing firm just eight weeks after new production processes were introduced.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

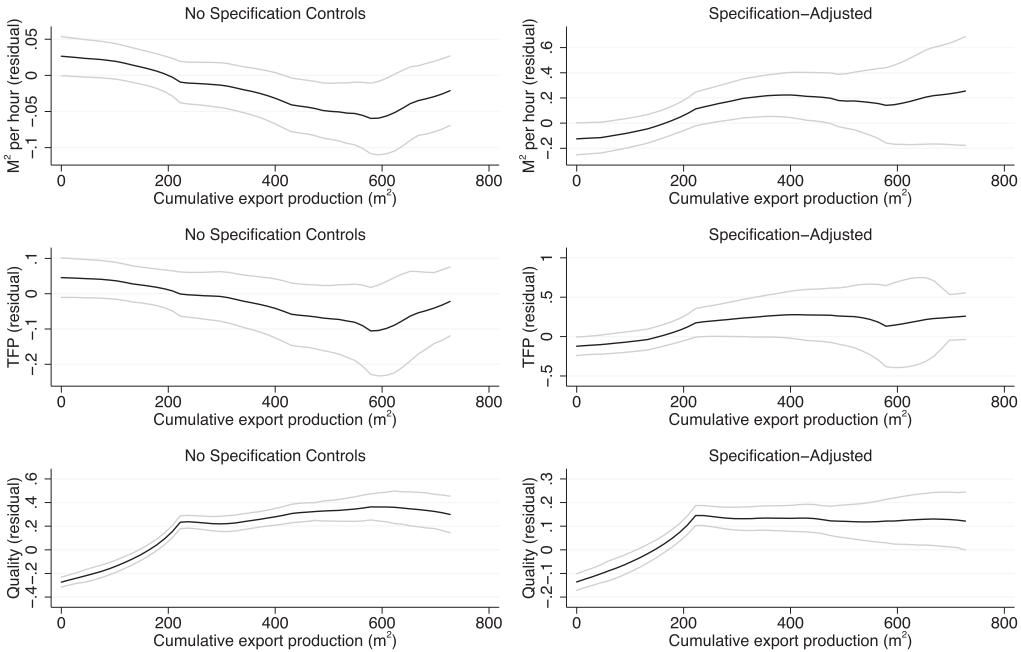

FIGURE V

Learning Curves for Takeup Firms (Step 3)

Figure plots learning curves obtained by regressing quality or productivity measures on firm and round fixed effects and then plotting a kernel-weighted local polynomial of the residuals against cumulative export production for takeup firms. Figure restricts attention to the range of cumulative exports achieved by Sample 2 firms.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

of a learning explanation. Online Appendix Figures F.2 and F.3 complement these results by showing similar patterns using four higher-frequency quality metrics collected by the intermediary for each batch of rugs delivered by each takeup firm.

VI.E. Step 4: Knowledge Transfers

The results in Steps 1–3 indicate that learning-by-exporting is present in our context. In this step, we distinguish between two types of learning-by-exporting discussed in the literature. The first is a learning-by-doing story where learning curves are particularly steep for the high-quality items demanded by foreigners and so the learning-by-doing is induced by exporting. The second is a story where actual knowledge is transferred between buyers, the intermediary and producers. Of course, we believe both are occurring, and this subsection simply provides evidence that some of the learning comes from knowledge transfers.

To measure knowledge flows, we tracked information flowing between buyers and Hamis, as well as between Hamis and the producing firms.

The data on flows between buyers and Hamis come from email correspondences Hamis shared with us. Here we provide several excerpts relating to various aspects of rug quality. In one correspondence, a foreign buyer complained that the rug was packed too tightly which results in wavy rugs: "Wrapping the kelims tightly and strongly leaves waving marks on them, so please roll kelims and wrap them softly to avoid waviness." On a separate occasion, the same buyer also noted that the edges of some carpets had frayed:

> We have a problem with our client. As you remember, this client asked for two carpets with fringes in the colour uni 2 and 3. Now after one and a half year using the carpets, the fringes crumble away, as you see on the pictures [reproduced in Online Appendix Figure F.6]. They will have two new pieces and give the whole problem to a lawyer. What to do?

These conversations suggest that buyers are passing along both information on how to manufacture high-quality rugs (e.g., packing that is not too tight) as well as information on what a high-quality product is (e.g., the importance of long-term durability). However, the correspondences are more suggestive in nature and are incomplete—we have no records of the information flows resulting from Hamis Carpets' export experiences prior to our study

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

(although Hamis reports having learned about both weaving techniques and quality control from previous interactions with foreign buyers).[32]

We have more detailed data on information flows between the intermediary and the firms. Hamis provided us with a log book of the visits made to each of the treatment firms as well as the subject discussed during that visit. In particular, we know the total number of conversations, their average length, and the topics discussed over the project period (Online Appendix Table F.7 presents summary statistics from this data set). The topics are categorized according to 10 of our 11 quality metrics (the intermediary did not discuss input quality since it provided the inputs). All takeup firms were visited at least 7 times, with the average firm visited 11 times. A visit lasted 28 minutes on average. They talked about issues related to design accuracy, the weight of the rug and the tightness of the warp thread on at least half of the occasions.

To find out more about the nature of these discussions, in August 2014 we asked firms whether their conversations along each of these dimensions involved the intermediary providing "information on techniques to improve quality" or just involved the intermediary "pointing out flaws." The data show that 89.4% of discussions were of the former type, discussions about production techniques to improve quality. The intermediary has provided us with multiple examples of the techniques passed on in these conversations. For example, the intermediary provided knowledge about the optimal way to weave the weft thread through the warp so as to achieve the correct firmness of the rug, about how to hold the weft thread to reduce waviness, and about how to maintain the integrity of the rug corners.

We can use these data to examine whether the knowledge imparted on these visits correlates with the improved performance of takeup firms. We match the data set of topics discussed during visits with each firm to the quality metrics recorded in the final survey round. This match allows us to test whether quality increased most along the particular quality dimensions discussed

32. For example, Hamis learned how to reduce the waviness of rugs by regularly adjusting the tenseness of the warp thread from a large furniture distributor near Bonn, Germany; and how to ensure consistency of designs over large orders by using measuring sticks and folding the rugs over to check for symmetry from a small artisan rug shop near Mannheim, Germany.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

with Hamis. To perform this test, we once more stack the quality (or specification-adjusted quality) measures, indexed by d, and run the following cross-sectional regression:

$$\begin{aligned} Quality_{id} &= \alpha_3 + \beta_3 Takeup_i \times \mathbf{1}[Talked\_About\_Dimension]_{id} \\ &\quad + \gamma_3 Quality_{id0} + \delta_i + \delta_d + \varepsilon_{id}. \end{aligned} \tag{10}$$

We include firm fixed effects δ_i so that we are comparing improvements in quality across the different dimensions d within the same firm. We also include quality metric fixed effects δ_d to control for different means across dimensions.[33] A significant β_3 coefficient is supportive of the presence of knowledge transfers as it implies that quality improves more along dimensions that are discussed than along dimensions that are not (and is inconsistent with a simple movement along the PPF, where quality would be independent of knowledge flows).

The results in Table XII support the hypothesis that knowledge is transferred from the intermediary to the firm. We find a positive and statistically significant association between changes in quality and whether the intermediary discussed that quality dimension with the firm, using either our standard quality metrics in Panel A, column (1) or our specification-adjusted ones in column (2). The size of the coefficient implies that quality levels improve by 16% more when the intermediary spoke to the firm about that dimension.[34] Panel A, columns (3) and (4) allow the β_3 coefficient to differ depending on whether the intermediary provided information on techniques or was just pointing out flaws. We find that both types of information are associated with improvements in quality, with the magnitudes of the two coefficients virtually identical. Finally Panel B repeats the analysis but includes additional interactions between $Takeup_i$ and each quality metric to allow for differential treatment effects by dimension. The

33. Note that we do not include controls for cumulative production or the takeup main effect since both vary only across firms and so are swept out by the firm fixed effects (as would any specification controls).

34. There is no evidence that firms achieve higher quality on the talked about dimension by reducing effort on other dimensions: there is a positive and significant coefficient on the interaction term if quality is regressed on $Takeup_i$ and $Takeup_i$ interacted with a dummy for whether Hamis discussed any quality dimension with them (with baseline controls in lieu of firm fixed effects). This result is available on request.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE XII
INFORMATION FLOWS AND QUALITY LEVELS (STEP 4)

	(1) Stacked quality metrics	(2) Specification-adjusted quality metrics	(3) Stacked quality metrics	(4) Specification-adjusted quality metrics
Panel A: Baseline				
Takeup$_i$ × *Talked About Dimension*$_{id}$	0.19**	0.16**		
	(0.08)	(0.07)		
Takeup$_i$ × *Information on Techniques for Dimension*$_{id}$			0.32***	0.33***
			(0.09)	(0.09)
Takeup$_i$ × *Pointed Out Mistakes in Dimension*$_{id}$			0.30***	0.29***
			(0.04)	(0.04)
Quality metric FEs	Yes	Yes	Yes	Yes
Takeup$_i$ × quality metric FEs	No	No	No	No
Firm FEs	Yes	Yes	Yes	Yes
Specification-adjusted quality metrics	No	Yes	No	Yes
R-squared	0.76	0.43	0.75	0.42
Observations	1,700	1,667	1,670	1,637

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

TABLE XII
CONTINUED

	(1) Stacked quality metrics	(2) Specification-adjusted quality metrics	(3) Stacked quality metrics	(4) Specification-adjusted quality metrics
Panel B: With "takeup × quality" FEs				
$Takeup_i \times Talked\ About\ Dimension_{id}$	0.15***	0.13**		
	(0.05)	(0.05)		
$Takeup_i \times Information\ on\ Techniques\ for\ Dimension_{id}$			0.16*	0.16
			(0.09)	(0.10)
$Takeup_i \times Pointed\ Out\ Mistakes\ in\ Dimension_{id}$			0.17**	0.15*
			(0.08)	(0.08)
Quality metric FEs	Yes	Yes	Yes	Yes
$Takeup_i$ × quality metric FEs	Yes	Yes	Yes	Yes
Firm FEs	Yes	Yes	Yes	Yes
Specification-adjusted quality metrics	No	Yes	No	Yes
R-squared	0.81	0.56	0.81	0.55
Observations	1,700	1,667	1,670	1,637

Notes. Table regresses stacked quality metrics on takeup indicator and its interaction with a dummy that takes the value 1 if the intermediary talked to the firm about that particular dimension of quality. Even-numbered columns use the specification-adjusted quality metrics described in the text. Columns (3) and (4) separate whether the discussion along that dimension was about technique or just pointing out mistakes. Regressions are run on a cross-section of firms and include baseline values, firm fixed effects, quality metric fixed effects, and quality metric fixed effects interacted with takeup. Panel B repeats the regressions in Panel A but also includes takeup interacted with quality metric fixed effects. Standard errors are clustered by firm. Significance: *.10; **.05; ***.01.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

coefficients in Panel B, columns (1) and (2) are almost identical to those in Panel A and highly significant. While the coefficients actually become larger than uninteracted coefficients when the type-of-information interactions are included in Panel B, columns (3) and (4), the higher standard errors mean that significance levels decline, with the technique interaction for specification-adjusted quality falling just below the 10% threshold.

We provide one additional piece of evidence that suggests our results are not driven by learning-by-doing alone. Under learning-by-doing, we would expect firms that were already producing high-quality rugs at baseline to see smaller treatment effects as they had less to learn.

This prediction is not borne out by the data: when we regress the stacked quality metrics on a treatment dummy, baseline quality and an interaction of the two, the interaction coefficient is insignificant.[35]

It is hard to completely dismiss the possibility that these discussions communicate what firms can get away with or the rug preferences of foreigners. However, the picture painted by these knowledge flows results is fully consistent with the quality lab results in Step 2. The examples above of superior weaving techniques discussed during these visits are all techniques that once acquired are costless to implement. Given that the majority of discussions are of this type, it is little surprise that treatment firms produced substantially higher quality domestic rugs in the quality lab despite having no incentive to do so.

VI.F. Step 5: Ruling Out Alternative Hypotheses

In this final step, we rule out alternative explanations that could explain the patterns in the data. There are three main competing hypotheses.

The first is that firms incur an adjustment cost while moving along the PPF, which could generate learning curves of the type we found in Step 3. While a reasonable story, adjustment costs alone cannot explain our findings in the other steps. For example, adjustment costs would not lead treatment firms to produce identical domestic rugs at higher quality many months after the intervention (Step 2).

35. We find a coefficient on the interaction of 0.05 (std. err. 0.04). As elsewhere, we include round and strata fixed effects.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

The second hypothesis is that we are simply picking up scale effects poorly captured by our production function. But we found no evidence that, relative to control firms, treatment firms increased scale by hiring more workers or expanding the number of looms (see Tables VI and VII). Although treatment firms did increase scale through more labor hours and longer production runs, it is implausible to attribute the Step 2 quality lab results to this greater scale because every firm had the same-length production run in the quality lab. Moreover, using the estimated relationship between quality and hours worked in the control group, the 5% increase in labor hours (from Tables VI and VII) would increase quality by less than one hundredth of the difference between treatment and control in the quality lab (from Table XI).[36]

A third closely related hypothesis is that the opportunity to export raised the returns to investments that raise output or quality but are not adequately captured by increases in our simple measures of capital and labor inputs. These investments could take the form of purchasing equipment, investing time in learning new techniques, or hiring consultants to teach new skills. If we do not account for these investments, we may spuriously conclude that there was learning-by-exporting.

Our data allow us to dismiss a simple investments hypothesis. First, we regularly surveyed firms about investments or costs incurred throughout the study. There is no quantitative (or qualitative) support indicating that treatment firms undertook any such investments. For example, no firm reports investing in a new loom or paying to repair existing looms over the duration of the sample. In addition we asked treatment firms about the extent to which they practiced weaving techniques, and none report ever practicing techniques.

A more complicated variant of the investments hypothesis would be that our intermediary provided a teacher or consultant to train treatment firms in weaving skills. If the intermediary deducted training costs from payments to the firm, this would be equivalent to an investment by the firm. However, we find no evidence of this type of payment: the price paid to firms is

36. Specifically, we regress the master artisan quality scores in the quality lab on log average monthly hours post baseline among the control group (controlling for baseline quality, baseline hours worked, and round and strata fixed effects), and apply the (insignificant) point estimate of 0.036 to the 5% increase in labor hours.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

uncorrelated with the number of hours the firm was visited by the intermediary.[37] Instead, the knowledge transfers occurring during these interactions appear to be flows of information that are not priced, exactly the type of information flows described in the classic learning-by-exporting literature (e.g., Clerides, Lach, and Tybout 1998).

A final variant of the investments hypothesis is that firms invest their time to raise the quality of their labor (i.e., trade off slower production now for higher returns from exporting once they can produce high-quality rugs with their improved labor). Inconsistent with this hypothesis is the fact that the benefits of the knowledge transfer, even in the domestic market, exceed the likely time costs incurred if higher quality could be purposefully learned.

To determine the value of improving quality for a firm that sells to the domestic market, we regress profits per hour for non-takeup firms (i.e., those selling to the domestic market) on our quality metrics.[38] Combining these estimates of the domestic returns to quality with the treatment effects estimated in the quality lab, the benefit for control firms to move to the quality levels achieved by treatment firms would be 9.96% higher profits on the domestic market. We assume that it takes five months to learn how to produce high quality (the approximate time taken to weave the 200 m^2 after which learning stops in the bottom row of Figure V).

Even if we assume that firms do not benefit at all from higher profits per hour until this learning period is over, profits would fall by only 5% in the learning period (taking the productivity drop between 0 and 200 m^2 in the first row of Figure V) and the firm would recover the investment in quality upgrading after eight months using an annualized discount rate of 10%. Also note that, using the wage the intermediary pays its employees who visit the firms and assuming all the time spent on firm visits is spent discussing techniques, the cost of providing this training is only

37. A regression among takeup firms of the log price received on log total hours of visits by the intermediary and specification controls gives a negative and insignificant coefficient of −0.06 (std. err. 0.06).

38. This is a regression of log profits per hour on the nine quality metrics recorded in our “Step 2” quality lab (as well as specification controls and round and strata fixed effects). The test that all the quality coefficients are jointly 0 is rejected at the 1.3% level. The regression is reported in Online Appendix Table F.8.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

LE103 ($16) compared to a lifetime net present value for firms of LE10,070 ($1,596). Hence, these calculations suggest that if firms were choosing whether or not to invest in learning the skills needed for high-skill production, they should have already chosen to make this investment to produce for the domestic market.

VII. Conclusion

This article conducts the first RCT that generates exogenous variation in the opportunity to export to understand the impacts of exporting on firm performance. The random variation, the detailed survey collection, and our quality lab allow us to make causal inferences about the impact of exporting and to identify the mechanisms through which improvements occur.

We find that profits for treatment firms increase 16–26% relative to control. This finding stands in contrast to many RCTs designed to alleviate supply-side constraints that have shown limited impacts on profits. Thus, our profit results suggest that demand-side constraints may be a critical barrier to firm growth in developing countries and can be mitigated through market access initiatives. The question of whether this market access program is cost effective and/or alleviates market failures is an interesting one which we leave for future work.

The rise in profits is driven by substantial quality upgrading accompanied by declines in output per hour. However, the quality upgrading we observe may or may not come about through learning-by-exporting—export-induced improvements in technical efficiency.

We provide five pieces of evidence that learning-by-exporting is occurring in our context. First, conditional on product specifications, we observe large improvements in both quality and productivity. Second, when asked to produce an identical domestic rug, treatment firms produce higher-quality rugs and do not take longer to do so. Third, we observe learning curves among the firms who took up the opportunity to export. Fourth, we document knowledge flowing between foreign buyers, the intermediary, and the producers, with quality increasing most along the specific dimensions that the knowledge pertained to. Fifth, we find no evidence that firms make monetary or time investments in upgrading, or pay, even implicitly, for the knowledge they receive.

Taken together, the evidence indicates that learning-by-exporting is present in our data and that the learning occurs, at

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

least in part, through information flows. Given that this learning is induced by demand for high-quality products from knowledgeable buyers in high-income countries, these changes would likely not have occurred as a result of increased market access to domestic markets.

As is the case in any analysis of a particular industry or location, we are cautious to generalize our findings too broadly. However, we believe that two features of this study—random assignment of export status and detailed surveys that allow us to unpack the changes occurring within the firms—contribute to the literature that studies the impacts of trade on the developing world.

Appendix: Measuring Unadjusted and Specification-Adjusted TFP

A key challenge in standard productivity analysis is the lack of firm-specific input and output prices which biases productivity estimates (see de Loecker and Goldberg 2014). We avoid these measurement issues because we observe output quantities. Moreover, since all firms produce a single product—handmade rugs—issues that arise with multiproduct firms and the divisibility of inputs are not relevant in this setting (de Loecker et al. 2016).

The first production function estimation does not control for rug specifications and hence provides our unadjusted TFP estimates. We estimate the following Cobb-Douglas production function:

$$\ln x_{it} = \lambda + \alpha_l \ln l_{it} + \alpha_k \ln k_{it} + a_{it}^{unadjusted} + \epsilon_{it}, \tag{11}$$

where x_{it} is the output (in m^2) of firm i in period t, l_{it} is total labor hours used, k_{it} is the number of active looms, and $a_{it}^{unadjusted}$ is the firm's unadjusted TFP. The error term captures unanticipated shocks as well as an omitted variable, the specifications of the rugs produced.

The second production function estimation controls for rug specifications and provides our specification-adjusted TFP estimate. We estimate

$$\ln x_{it} = \lambda + \gamma_l \ln l_{it} + \gamma_k \ln k_{it} + \mathbf{Z}_{it}^{'}\Gamma + a_{it}^{adjusted} + \nu_{it}, \tag{12}$$

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

where $a_{it}^{adjusted}$ is the firm's adjusted TFP and the vector $\mathbf{Z}_{it}$ includes six rug specifications: rug difficulty, thread count, thread type, number of colors, market segment, and narrow product type.[39] The error term now only captures unanticipated shocks (and, of course, measurement error).

We do not include the costs incurred to find foreign buyers in the production function for physical output. Common to the learning-by-exporting literature, we view any productivity gains from exporting as unanticipated by the firm. That is, firms make exporting decisions without anticipating that they would become more efficient producers. The evidence in Section VI supports this claim by showing that productivity improvements come, at least in part, from knowledge transfers that the firms do not pay for. In this case, the costs incurred to find foreign buyers should not be included in the production function for physical output under a reasonable definition of TFP.[40]

Although data on quantities and specifications deal with measurement concerns, there is still potential simultaneity bias since TFP is observed by the firm but not by us. We follow the control function approach (Olley and Pakes 1996) and assume capital is subject to adjustment costs, labor is a flexible input, and we use warp thread quantity as the proxy. We exploit the experimental design by estimating the production function using only control firms, avoiding the need for assumptions on how treatment changes the TFP process.[41] 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. For the production function in equation (11), we obtain $\alpha_l = 0.74$

39. In the Sample 1 baseline, we did not record the market segment or rug difficulty. We replace these missing values with the corresponding values from the subsequent survey round.

40. We also note that, unanticipated or not, to our knowledge resources associated with generating export orders have never been explicitly included in any production function estimation. For example, de Loecker (2007, 2013) includes an export dummy in the productivity law of motion, rather than as a separate factor of production.

41. See de Loecker (2013) for an extensive discussion of this point. To check the stability of the production function coefficients, we ran an OLS estimation of equation (12) on baseline and endline data, and allowed the α_l and α_k coefficients to vary by round and treatment/control group; we find no statistical difference between the coefficients. This suggests that while the level of TFP changed among treatment firms, their production function did not.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

(std. err. of 0.53) and $\alpha_k = 0.30$ (std. err. of 0.20).[42] For the production function in equation (12), we obtain $\alpha_l = 1.11$ (std. err. of 0.30) and $\alpha_k = 0.19$ (std. err. of 0.11). We cannot reject the null of constant returns to scale in either case.

Having estimated the coefficients, we compute $a_{it}^{unadjusted} = \ln x_{it} - \hat{\alpha}_l l_{it} - \hat{\alpha}_k k_{it}$ and $a_{it}^{adjusted} = \ln x_{it} - \hat{\gamma}_l l_{it} - \hat{\gamma}_k k_{it} - \mathbf{Z}_{it}^{'}\hat{\Gamma}$, where hats denote estimated parameters. Note that the methodology assumes that the parameters of the production function are the same for treatment and control firms. This is a reasonable assumption because all firms produce rugs using the same technology in all periods. Since all firms produce a narrowly defined product, the assumption is weaker than existing work that typically assumes identical parameters for all firms within a two- or four-digit industry classification.

MIT
Columbia Business School
University of Illinois at Urbana-Champaign

Supplementary Material

An Online Appendix for this article can be found at *The Quarterly Journal of Economics* online.

REFERENCES

Ahn, JaeBin, Amit K. Khandelwal, and Shang-Jin Wei, "The Role of Intermediaries in Facilitating Trade," *Journal of International Economics*, 84 (2011), 73–85.
Allen, Treb, "Information Frictions in Trade," *Econometrica*, 82 (2014), 2041–2083.
Alvarez, Fernando E., Francisco J. Buera, and Robert E. Lucas Jr., "Idea Flows, Economic Growth, and Trade," NBER Working Paper 19667, 2013.
Arkolakis, Costas, Arnaud Costinot, and Andres Rodriguez-Clare, "New Trade Models, Same Old Gains?," *American Economic Review*, 102 (2012), 94–130.
Baltagi, Badi H., and Dong Li, "Series Estimation of Partially Linear Panel Data Models with Fixed Effects," *Annals of Economics and Finance*, 3 (2002), 103–116.
Banerjee, Abhijit V., "Microcredit under the Microscope: What Have We Learned in the Past Two Decades, and What Do We Need to Know?," *Annual Review of Economics*, 5 (2013), 487–519.
Bastos, Paulo, Joana Silva, and Eric Verhoogen, "Export Destinations and Input Prices," Unpublished manuscript, 2014.
Bernard, Andrew B., and B. Jensen, "Exceptional Exporter Performance: Cause, Effect, or Both?," *Journal of International Economics*, 47 (1999), 1–25.

42. For comparison, the OLS of equation (11) gives $\alpha_l = 0.82$ (std. err. of 0.12) and $\alpha_k = 0.33$ (std. err. of 0.11).

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David McKenzie, and John Roberts, "Does Management Matter? Evidence from India," *Quarterly Journal of Economics*, 128 (2013), 1–51.
Bloom, Nicholas, and John Van Reenen, "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 122 (2007), 1351–1408.
——, "Why Do Management Practices Differ across Firms and Countries?," *Journal of Economic Perspectives*, 24 (2010), 203–224.
Brambilla, Irene, Daniel Lederman, and Guido Porto, "Exports, Export Destinations, and Skills," *American Economic Review*, 102 (2012), 3406–3438.
Clerides, Sofronis K., Saul Lach, and James R. Tybout, "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco," *Quarterly Journal of Economics*, 113 (1998), 903–947.
Crozet, Matthieu, Keith Head, and Thierry Mayer, "Quality Sorting and Trade: Firm-level Evidence for French Wine," *Review of Economic Studies*, 79 (2012), 609–644.
de Loecker, Jan, "Do Exports Generate Higher Productivity? Evidence from Slovenia," *Journal of International Economics*, 73 (2007), 69–98.
——, "Detecting Learning by Exporting," *American Economic Journal: Microeconomics*, 5 (2013), 1–21.
de Loecker, Jan, and Pinelopi Goldberg, "Firm Performance in a Global Market," *Annual Review of Economics*, 6 (2014), 201–227.
de Loecker, Jan, Pinelopi Goldberg, Amit K. Khandelwal, and Nina Pavcnik, "Prices, Markups and Trade Reform," *Econometrica*, 84 (2016), 445–510.
de Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 123 (2008), 1329–1372.
——, "Measuring Microenterprise Profits: Must We Ask How the Sausage is Made?," *Journal of Development Economics*, 88 (2009), 19–31.
——, "Wage Subsidies for Microenterprises," *American Economic Review*, 100 (2010), 614–618.
——, "Business Training and Female Enterprise Start-up, Growth, and Dynamics: Experimental Evidence from Sri Lanka," *Journal of Development Economics*, 106 (2014), 199–210.
Eaton, Jonathan, Marcela Eslava, David Jinkins, C. J. Krizan, and James Tybout, "A Search and Learning Model of Export Dynamics," Technical report, Penn State, 2013.
Feenstra, Robert C., and John Romalis, "International Prices and Endogenous Quality," *Quarterly Journal of Economics*, 129 (2014), 477–527.
Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 98 (2008), 394–425.
Hall, Robert, and Charles Jones, "Why Do Some Countries Produce so Much More Output per Worker than Others?," *Quarterly Journal of Economics*, 114 (1999), 83–116.
Hallak, Juan Carlos, "Product Quality and the Direction of Trade," *Journal of International Economics*, 68 (2006), 238–265.
——, "A Product-Quality View of the Linder Hypothesis," *Review of Economics and Statistics*, 92 (2010), 453–466.
Hallak, Juan Carlos, and Peter K. Schott, "Estimating Cross-Country Differences in Product Quality," *Quarterly Journal of Economics*, 126 (2011), 417–474.
Hallak, Juan Carlos, and Jagadeesh Sivadasan, "Product and Process Productivity: Implications for Quality Choice and Conditional Exporter Premia," *Journal of International Economics*, 91 (2013), 53–67.
Harrison, Ann, and Andres Rodriguez-Clare, "Trade, Foreign Investment, and Industrial Policy for Developing Countries," vol. 5 of Handbook of Development Economics (Amsterdam: Elsevier, 2010), 4039–4214.
Keller, Wolfgang, "International Technology Diffusion," *Journal of Economic Literature*, 42 (2004), 752–782.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019

Khandelwal, Amit, “The Long and Short (of) Quality Ladders,” *Review of Economic Studies*, 77 (2010), 1450–1476.
Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, “Experimental Analysis of Neighborhood Effects,” *Econometrica*, 75 (2007), 83–119.
Krugman, Paul R., “Increasing Returns, Monopolistic Competition, and International Trade,” *Journal of International Economics*, 9 (1979), 469–479.
Kugler, Maurice, and Eric Verhoogen, “Prices, Plant Size, and Product Quality,” *Review of Economic Studies*, 79 (2012), 307–339.
Lederman, Daniel, Marcelo Olarreaga, and Lucy Payton, “Export Promotion Agencies: Do they Work?,” *Journal of Development Economics*, 91 (2010), 257–265.
Levitt, Steven D., John A. List, and Chad Syverson, “Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant,” *Journal of Political Economy*, 121 (2013), 643–681.
Manova, Kalina, and Zhiwei Zhang, “Export Prices across Firms and Destinations,” *Quarterly Journal of Economics*, 127 (2012), 379–436.
Marin, Alvaro Garcia, and Nico Voigtlander, “Exporting and Plant-Level Efficiency Gains: It’s in the Measure,” NBER Working Paper 19033, 2013.
McKenzie, David, “Beyond Baseline and Follow-up: The Case for More T in Experiments,” *Journal of Development Economics*, 99 (2012), 210–221.
McKenzie, David, and Christopher Woodruff, “What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?,” *World Bank Research Observer*, 29 (2013), 48–82.
Melitz, Marc J., “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity,” *Econometrica*, 71 (2003), 1695–1725.
Olley, G. Steven, and Ariel Pakes, “The Dynamics of Productivity in the Telecommunications Equipment Industry,” *Econometrica*, 64 (1996), 1263–1297.
Park, Albert, Dean Yang, Xinzheng Shi Shi, and Yuan Jiang, “Exporting and Firm Performance: Chinese Exporters and the Asian Financial Crisis,” *Review of Economics and Statistics*, 92 (2010), 822–842.
Schott, Peter K., “Across-Product versus Within-Product Specialization in International Trade,” *Quarterly Journal of Economics*, 119 (2004), 647–678.
Syverson, Chad, “What Determines Productivity?,” *Journal of Economic Literature*, 49 (2011), 326–365.
Verhoogen, Eric A., “Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector,” *Quarterly Journal of Economics*, 123 (2008), 489–530.
Wagner, Joachim, “Exports and Productivity: A Survey of the Evidence from Firm-level Data,” *World Economy*, 30 (2007), 60–82.
Wooldridge, Jeffrey M., “On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables,” *Economics Letters*, 104 (2009), 112–114.

Downloaded from https://academic.oup.com/qje/article-abstract/132/2/551/3002609 by guest on 01 August 2019