NBER WORKING PAPER SERIES

EXPORTING AND FIRM PERFORMANCE:
EVIDENCE FROM A RANDOMIZED TRIAL

David Atkin
Amit K. Khandelwal
Adam Osman

Working Paper 20690
http://www.nber.org/papers/w20690

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2014

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 toArtisans including Lisa Smith, Shereen Shirazy, Mary Cockram, Maud Mabika, and Bill Kruvant at Creative Learning. We are grateful to Josephine Gantois for her excellent research assistance. We also wish to thank Nicholas Bloom, Juan Carlos Hallak, David McKenzie, Nina Pavcnik, Jim Tybout, Eric Verhoogen, ChristopherWoodruff, 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 Chazen Institute at Columbia Business School. Yale IRB Approval #1001006247 (Atkin and Osman were at Yale University during the experiment), Columbia IRB Approval #AAAE9678. This RCT was registered in the American Economic Association Registry for randomized control trials under Trial number AEARCTR-0000069. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Exporting and Firm Performance: Evidence from a Randomized Trial
David Atkin, Amit K. Khandelwal, and Adam Osman
NBER Working Paper No. 20690
November 2014, Revised April 2015
JEL No. F10

ABSTRACT

We conduct a randomized control trial 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 profits and productivity. Treatment firms report 15-25 percent 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 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.

David Atkin
Department of Economics
University of California, Los Angeles
8283 Bunche Hall
315 Portola Plaza
Los Angeles, CA 90095
and NBER
atkin@econ.ucla.edu

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

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

1 Introduction

There are large differences in productivity across countries (Hall and Jones 1999, Bloom and Van Reenen 2007). The belief that access to high-income country 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 WTO's Aid-for-Trade Initiative secured $48 billion in annual commitments to help developing countries overcome "trade-related constraints". Central to achieving this goal is the idea that exporting improves the productivity of firms, a mechanism referred to as learning-by-exporting (Clerides et al. 1998, de Loecker 2007, Harrison and Rodriguez-Clare 2010).

Despite the pervasiveness of these initiatives, there is still an ongoing debate as to whether exporting has a causal impact on productivity. Moreover, if productivity does change, the mechanisms are not well understood. There are two central challenges in identifying potential causal effects of exporting. First, more productive firms select into exporting (see the survey by Melitz and Redding 2014). This selection has plagued empirical attempts to identify the impact of exporting on firm performance 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. Instead, the literature uses residual-based measures, such as revenue-based TFP, which also capture changes in product specifications, product mix, investments, markups and input costs (de Loecker and Goldberg 2014). If trade causes firms to upgrade along these dimensions, rising TFP measures may simply reflect movements along the production possibility frontier (PPF), rather than outward shifts of the PPF.

This paper conducts a randomized control trial (RCT) on rug manufacturers in Egypt to examine how exporting affects profits and productivity. To be clear, the goal of this paper 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 (a question that would require an entirely different experimental design). Instead, we use experimental variation to uncover economic primitives, in particular whether, and how, productivity evolves in response to exporting. To our knowledge, this is the first attempt to generate exogenous firm-level variation in the opportunity to export.

The random assignment into exporting directly addresses the first challenge: selection of firms into exporting. Specifically, we provided a subset of firms the opportunity to export handmade carpets to high-income markets. To provide this opportunity, we partnered with a US-based non-governmental 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, firms with 1 to 4 employees, located in Fowa, Egypt. A random subsample of these firms was provided with an initial opportunity to fill the orders by producing 110m^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 non-exporting firms with the opportunity to export 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 (who received the opportunity to export) and control firms (who received no such opportunity). Focusing the analysis on a single industry provides several advantages that we exploit in the analysis: the production technology is homogenous across firms, quality metrics are well-defined and codifiable, and physical productivity can be accurately measured. For example, the literature typically relies on prices and input costs to infer product quality (e.g., Schott 2004 or Hallak 2006). In contrast, our production-line level data allow us to record detailed specifications for the rugs being produced at the time of each survey round. These specifications include product categories (analogous to product codes used in the literature, but at a much finer level) as well as attributes, such as thread count, that a buyer chooses when they place their order. This level of detail allows us to construct measures of physical output productivity that are comparable across firms. Moreover, we complement these specifications with 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 characteristics 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. Together, these data allow us to address directly the measurement challenges in assessing how exporting affects firm performance.

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 15-25 percent higher profits relative to control firms. The substantial increase in profits is perhaps not surprising, but is interesting given the more moderate profit impacts the literature has found when exploring supply-side interventions such as credit access (Banerjee, 2013). It is also suggestive that the distributional consequences of trade may come in part from heterogeneity in market access.

The primary focus of this paper is to understand the mechanisms driving the profit increases. Treatment firms report increases in output prices, input prices and labor hours. However, despite these increases, 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 are slower to produce because of their difficulty. Indeed, our data confirm that output from treatment firms receive significantly higher scores along virtually every quality dimension. At the same time, physical productivity (not adjusted for rug specifications) falls by 24 percent among treatment firms.

A simple theoretical framework shows that these findings are consistent with two distinct mechanisms that have not been disentangled in the literature to date. We posit that the productivity 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 who 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 rugs of different quality levels and the opportunity to export 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, or learning-by-exporting as it is referred to in the literature. Learning-by-exporting includes *both* transfers of knowledge from buyers to producers, *and* 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 for a given set of specifications) or raising χ_q (producing higher quality conditional on specifications). When these increases in efficiency are biased towards the production of high-quality rugs, both rug quality and profits will rise. Of course, these two mechanisms are not mutually exclusive, but the presence of learning-by-exporting is potentially important for both theory and policy because it implies an improvement in efficiency. Unlike previous studies, the random assignment of export orders and our detailed data allows us to distinguish the two mechanisms.

We present five pieces of evidence to show that the improvements in performance come, at least in part, through the learning-by-exporting mechanism. The first is that both quality and productivity *rise* after adjusting for product specifications (recall that without conditioning, productivity falls). If firms only moved along the PPF, quality and productivity would remain constant after adjusting. Second, at the endline, we asked all firms in our sample to manufacture an *identical* domestic rug using identical inputs and a common loom in a workshop that we leased (a "quality lab"). The rugs that treatment firms produce 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 should immediately jump and then stay fixed), we document learning curves. Rug quality increases with cumulative export production, and similarly, unadjusted productivity initially drops upon exporting and then gradually rises over time (while adjusted productivity smoothly increases over time). Fourth, we draw on correspondences between foreign buyers and the intermediary, as well as a log book of discussions between the intermediary and producers, to document that our results come, in part, from knowledge flows (information that would be irrelevant if firms were only moving along the PPF). In particular, we show that treatment firms improve quality most along the particular quality dimensions that are discussed during meetings between the intermediary and the producer. This suggests that the improvements in efficiency occur, at least in part, through knowledge transfer from buyers. Fifth, we rule out adjustment cost or 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.

While not the central focus of our paper, before concluding we touch on the issue of market failures. In particular, if the knowledge generated by exporting is truly valuable, why does it not spillover to control firms and eliminate our treatment effects? We first establish that, even in the domestic market, the knowledge has a return that substantially exceeds the costs of providing it. Despite this, we find no evidence that the knowledge spilled over to the control firms. These findings suggest that firms either did not know this knowledge existed or were not able to purchase it, and points to a failure in the market for information similar to that discussed in Bloom et al. (2013).

For our results to have external validity, we must assume that productivity changes through learning-by-exporting in similar ways in other settings. For a variety of reasons, we believe this to be true. First, many authors have noted that a key benefit of exporting from developing to developed countries is the exposure to sophisticated buyers who have a stronger preference for higher-quality products than local buyers (e.g., Harrison and Rodriguez-Clare (2010), Verhoogen (2008), Park et al. (2010) and Artopoulos et al. (2013)).[1] Second, the process of exporting via an intermediary is typical not just for Egyptian rugs but for firms in many industries and countries: World Bank Enterprise Surveys reveal that 36 percent of exporters use an intermediary (and 62 percent for exporting firms with 5 or fewer employees). Third, the rug industry itself is also an important sector in other developing countries as we discuss in Section 2.1. All that said, the firms in our sample are small, typically having only one employee, and production is not automated so that the scope for technology upgrading may be limited. 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 basic technology improves statistical power (Bloom et al., 2013). Ultimately, as is the case for any industry or country study, be it the Taiwanese electronics industry (Aw et al., 2011) or homogeneous products in the U.S. (Foster et al., 2008), the external validity of our results is an empirical question, and the novel methodology we propose to identify learning-by-exporting can be applied to other settings.

Our results relate to a number of papers that span the trade and development literatures. Most directly, we contribute to a voluminous that seeks to identify the existence of learning-by-exporting. The evidence from these studies is mixed. For example, Clerides et al. (1998) and Bernard and Jensen (1999) find no support for the hypothesis and conclude that firms self-select into export markets. In contrast, several papers that use different techniques to deal with selection (e.g., matching estimators or instrumental variables) find some support for the hypothesis.[2] One implication of these studies is that in order to detect any potential learning-by-exporting, one must directly confront selection. A second implication is that even if one is convinced by the reduced-form approaches to deal with selection, data limitations prevent us from understanding

[1] Conversely, given the limited number of high-income consumers in a country like Egypt, we would not expect the quality upgrading and associated productivity improvements we find had we just expanded domestic market access.

[2] See de Loecker (2007), Park et al. (2010), Marin and Voigtlander (2013) and de Loecker (2013). Keller (2004), and Wagner (2007) and Harrison and Rodriguez-Clare (2010) survey the literature.

if measured productivity changes actually reflect outward shifts in the PPF or simply movements along the PPF. We contribute to this literature by directly confronting selection through random assignment; and directly confronting 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 results also relate to the literature on quality upgrading. Studies using country- or product-level data show that export quality positively co-varies with destination income-per-capita (Schott 2004, Hallak 2006 and Hallak 2010); and firm-level studies suggest that quality upgrading is paramount for export success, especially for developing countries.[3] Unlike much of this literature that must infer quality from price data or certifications, or through structural models where quality is inferred from prices and quantities, we collect direct measures of quality.[4] In addition to our randomization methodology and the comparatively rich survey data, we 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 et al., 2008, 2010, 2014 and 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 paper is organized as follows. Section 2 describes the research setting. Section 3 explains our experimental intervention and introduces the data. Section 4 examines the impact on profits and Section 5 decomposes the profit changes. Section 6 presents a theoretical framework that then guides our steps to detecting learning-by-exporting. Section 7 discusses spillovers and potential failures in the market for information. Section 8 concludes.

2 Research Setting

We first discuss the carpet industry in Fowa and why we chose this industry and location. We then describe the production technology for carpets. Finally, we discuss the process through which we generated the export orders necessary to carry out the experiment.

2.1 The Industry and the Location

In order to carry out a randomized evaluation of the impact of exporting, we sought out private-sector, governmental and non-governmental organizations involved in market access initiatives. In October 2009, we entered conversations with Aid to Artisans (ATA), a U.S.-based NGO with a mission to create economic opportunities for small-scale producers of handmade products

[3] For example, see Verhoogen (2008), Manova and Zhang (2012), Crozet et al. (2012), Brambilla et al. (2012), Hallak and Sivadasan (2013), and Bastos et al. (2014). The one exception is Marin and Voigtlander (2013), who find that rather than quality rising, marginal costs decline because Columbian firms make investments to lower marginal costs of production at the same time as they enter export markets.

[4] Papers that infer quality using structural approaches include Khandelwal (2010), Hallak and Schott (2011), and Feenstra and Romalis (2014). Crozet et al. (2012) is an exception that uses wine ratings as a measure of wine quality.

around the world. They had recently acquired USAID funding for a market access facilitation 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 OECD markets. First, ATA explores the country in question for products that would both appeal to high-income OECD consumers 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 that can 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 provides some training to the intermediary and then works closely with it to both produce appealing products and to market them. To produce appealing products, ATA draws on its experience in the handcrafts industry and will occasionally pay for design consultants. In terms of marketing the products, ATA prominently displays the products at major trade shows, for example the New York International Gift Fair (NYIGF), as well as drawing on its extensive network of contacts in the industry.

Working through a lead intermediary firm, rather than matching individual producers directly with foreign buyers, is an important aspect of the business model. The lead intermediary aggregates orders from many small producers. The ultimate objective is to foster self-sustaining relationships whereby ATA can eventually exit the sector. The hope is that the intermediary maintains, or preferably expands, its clients once ATA departs.[5]

This process through which exporting relationships emerge is not uncommon in other settings. Ahn et al. (2011) show that small-scale firms will use intermediaries to export in order to avoid large fixed costs associated with directly exporting. World Bank Enterprise Data record direct and indirect export activity of firms across many countries. Among manufacturing firms, 36 percent of exporters use an intermediary with this number rising to 62 percent when we restrict attention to firms with five or fewer employees to facilitate comparison with our context.[6] In our setting, the lead intermediary fulfills the role of aggregating orders and spreading the fixed costs of exporting across many small-scale producers. ATA acts as another middleman in this process, facilitating connections between domestic intermediaries and foreign buyers.

Alongside ATA, we searched for viable Egyptian products for more than a year before identifying handmade carpets from Fowa as having potential. Fowa is a peri-urban town located two hours southeast of Alexandria. The town has a population of 65,000 and lies in the governorate of Kafr El-Sheikh which has an average income per capita of \$3,600 (PPP-adjusted), well below the national average of \$6,500 (PPP-adjusted). Fowa is well known for its carpet cluster which contains hundreds of small firms that manufacture handmade textile products using wooden looms.

[5] For example, at the 2010 NYIGF, which drew 35,000 attendees, we met intermediaries from developing countries who were linked to Western markets by ATA and had now graduated to their own independent display at the fair.

[6] The data suggest that exporting via intermediaries is particularly common in the rug industry. For example, Chinese customs indicate that 37 percent of its exports in HS Code 580500 ("hand-woven tapestries") went through intermediaries compared to 20 percent of overall exports (Ahn et al. 2011). The need for such intermediation between suppliers and buyers has also been noted by Rauch (1999) and Feenstra and Hanson (2004).

These firms typically employ between 1 and 4 employees and produce flat-weave rugs, a product in which Egypt has a strong historical reputation.

Both the handmade craft industry and the rug industry are large and important sources of employment in developing economies. Global handmade craft production was estimated at $23.2 billion in 2005, while world production of carpets and rugs totaled $32 billion in 2008 (UNCTAD Creative Economy Report 2010). Egypt is the 11th largest producer of carpets and rugs with a total production at $734 million (36 percent of Egypt's total textile sector and 1.3 percent of total manufacturing output).[7] More than 17,000 people work in the carpets and rugs industry in Egypt, representing 7 percent of world employment in this industry and 1.7 percent of total manufacturing employment in Egypt (UNIDO 2013). Egypt has a (revealed) comparative advantage in this sector: in 2013 Egypt's exports in HS58 (special woven fabrics, tufted textiles, lace) constituted 0.6 percent percent of its exports, three times Egypt's share of world exports. These exports predominantly go to the U.S. (38 percent) and Western Europe (31 percent).

In November 2010 we identified a local intermediary, Hamis Carpets, as a potential partner for the program. Hamis is the largest intermediary in Fowa and accounts for around 20 percent of the market. At the time, Hamis earned 70 percent of its sales in the domestic market, mostly selling to distributors and retailers in Cairo, Alexandria and Luxor. ATA believed that, together with Hamis, they could generate additional orders from overseas buyers and fill these order by forming new relationships with small-scale producers in Fowa. ATA brought the CEO of Hamis to the US for a training course, provided marketing support and insisted that Hamis agree to the protocols of our experiment (described below). The marketing support included displaying the new products at various international gift fairs and at smaller events in Cairo. ATA also marketed Hamis' rugs to high-income OECD retailers through a US-based rug importer.[8]

2.2 Production Technology

A major benefit of our focus on the rug industry is that the production technology is identical across our sample firms. This both facilitates data collection (we can tailor the surveys to ask specific questions about production) and makes it possible to cleanly identify changes in the production function. In this section, we describe the technology in detail.

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. Firms self-identify as specialists in one of four flat-weave rug types: duble, tups, kasaees and goublan.[9] Duble rugs are the main rug

[7] Statistics from Euromonitor International Passport Database, Egypt national statistics, UN and OECD.

[8] ATA's USAID grant expired and in September 2012 it formally ended its involvement in this project and closed its Cairo office. However, Hamis Carpets agreed to continue participating in the research experiment after ATA exited. Hamis Carpets had several incentives to do so. First, we sponsored the CEO's visit to the NYIGF in January 2013. Second, there was one instance in which we provided a quarter of the capital ($7,000) to finance a relatively large sample order for a new client which was ultimately unsuccessful. Third, we provided $500 a month to offset costs of participating in the experiment (such as conducting rug quality surveys and filling out order books). Finally, the CEO is an active member of the Fowa weaving community and believes that showing how exporting improves the livelihoods of the local population will be valuable in promoting the sector.

[9] Duble and tups rugs are the most common types; kasaees rugs are woven from rags and are the cheapest type; goublan rugs are the most expensive type and are works of art used as wall hangings. See Appendix Figure B.1 for pictures.

type that Hamis Carpets, our intermediary, sells. As we explain later, our export orders ended up being almost exclusively for this type of rug.

The process of producing rugs is standardized across firms. The elements of the production technology are marked in Figure 1. The rugs are made on a large wooden foot-treadle loom. The width of the loom determines the maximum width of a rug. Rugs can be made of any length. The *warp thread* is the wool or cotton thread that spans the entire length of the rug and must be attached to the loom before rugs can be weaved. These threads cannot be seen on the final rug but are necessary to hold the rug together. The warp threads are kept in place using a reed which resembles a very large comb. The *weft thread* (typically made from wool) is the visible thread on the rug and is weaved between these warp threads using a shuttle. A foot-operated heddle is used to raise every alternate warp thread allowing the weaver to more quickly weave the weft threads between the warp threads. The weaver changes out the weft thread as he weaves based on the needs of the design until the rug is complete. At that point he cuts off the completed rug and continues to utilize the remaining warp thread until the production run of that particular type of rug is finished.

The average duble rug destined for domestic markets is sold by firms for LE42.5, and requires 5.9 hours of labor, per m^2.[10] After accounting for input costs, hourly wages are roughly LE3.[11]

There are several dimensions through which quality can vary across firms within rug types. First, the quality of the input thread varies across rugs. This dimension is determined primarily by the specifications of the rug ordered by the buyer with thicker rugs or different styles requiring different types and amounts of wool or cotton. There are also many dimensions of quality that depend primarily on weaving technique. For example, how flat the rug lies on a hard surface is determined by how well the warp and weft thread are installed on the loom. Similarly, other quality measures that depend on weaving technique include how well defined the corners are, how accurately the design was followed, and whether the rug adheres to the desired size specifications. Thus, manufacturing higher-quality rugs takes not only more labor inputs but also more skill. As discussed below, we have collected both rug specifications and these quality metrics for all firms.

2.3 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 OECD countries. Generating sustained orders was not guaranteed. The textile market is competitive and conversations with ATA's staff revealed that only 1 in 7 matches lead beyond trial orders. This is consistent with Eaton et al. (2013) who estimate that only 1 in 5 potential importer-exporter matches result in a successful business relationship.

ATA would typically introduce Hamis to foreign importers or retailers. Hamis and the foreign buyer would discuss pricing and delivery time, as well as product specification (design, colors, materials etc; see Figure 2 for an example of how these specifications are formalized after these dis-

[10] As discussed below, there are two baseline surveys that were run in July 2011 and February 2013. The exchange rate on July 1, 2011 was 5.94 Egyptian pounds (LE) to 1 U.S. dollar. The exchange rate on February 1, 2013 was LE 6.68. We will apply an average exchange rate of 6.31 unless otherwise noted.

[11] Tups and goublan rugs are three times more expensive but require 6-8 times more hours per m^2, while kasaees rugs are about one-fifth the price and take one-fifth the time.

cussions). Hamis would then organize the production of sample orders, either from its in-house weavers or ordered from one of the treatment firms in our sample.[12] This process can be costly, particularly in terms of time, as Hamis and the potential buyer iterate on design patterns, color schemes, technical aspects of quality, and price.

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. But as Figure B.1 illustrates, the style of goublan rugs is unlikely to appeal to high-income OECD buyers. Instead, it appears that high-income OECD buyers prefer "modern" designs, as illustrated in Figure 3. (The right-most rug in this figure is produced by one of our sample firms and retails for $1,400 in a high-end furniture store in the United States.)

After one-and-a-half years of searching, in June 2012, Hamis Carpets secured its first large export order (3,640m^2) from a German buyer and as of June 2014, its major buyers continue to place large, regular orders. Figure 4 reports that cumulative export production since December 2010 have totaled 33,227m^2. Our records indicate that cumulative payments to the producers has totaled LE982,351 ($155,682). As described in the next section, these orders were entirely sourced from our treatment firms, which forms the basis of our experiment.

3 The Experiment

3.1 Experimental Design

In July 2011, we compiled a list of firms in Fowa who had fewer than 5 employees, worked on their own account (meaning that they bought their own inputs), and had never previously worked with Hamis. We used an Egypt-based NGO to locate these firms since there is no census of carpet manufacturers in Fowa (all firms in our sample are informal) and many firms are located within homes making them difficult to find. These firms specialize in one of the four rug types described above, and we stratified the sample both on the type of rug produced and the loom size. We stratified on these two particular dimensions because they determine the types of order a particular firm can fill and we were concerned that ATA and Hamis would not be able to secure export orders for every rug type (which turned out to be the case).[13] For reasons that will be come clear momentarily, we refer to these 303 firms as "Sample 1". The first two rows of columns 1-4 of Table 1 show the total number of firms by rug type and treatment status for Sample 1.[14]

We designed the following export-market access intervention. Hamis Carpets (with ATA's assistance) marketed rugs to overseas buyers, and once export orders were secured we divided the order and allocated an *initial* amount to each of the producers in our treatment group. The treatment firms were visited by our survey team and a representative of Hamis carpets and offered the opportunity to fill the order. More precisely, Hamis Carpets showed them the rug design, ex-

[12]Throughout the project, Hamis carpets has employed a small number of workers who work on its premises producing samples and orders outside this research project.

[13]The loom size determines the maximum width of a rug a firm can manufacture, a specification chosen by the buyer.

[14]We randomized at the rug-type and loom-size level and some strata were uneven leading to 149 treatment firms out of the sample of 303 firms.

plained that the carpet would be exported to high-income OECD markets, and offered them an order of 110m^2 which translates to about 11 weeks of work. The 110m^2 was chosen as a balance between a reasonable size order and the ability 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 (prices we analyze in detail below). If the firm accepted, Hamis delivered the input thread and the correctly sized reed and heddle to ensure all rug orders were consistent across producers. At the same time, as is typical in many buyer-producer relationships, Hamis would discuss the technical aspects of the specific rug order and answer any questions the firm may have. Firms would deliver rugs to Hamis with payment upon delivery.

As further export orders were generated, Hamis continued to place them with the treatment firms. Just as in any arms-length transaction, after the initial order amounts were offered, Hamis was not bound to continue to make subsequent purchases from any particular treatment firm if the quality was below par or the previous rugs were not delivered on time. In other words, the experiment protocol simply forced Hamis to offer an *initial* order to the treatment firm. Hamis was not allowed to allocate any orders to control firms and we maintained a project coordinator and survey team in Fowa to ensure that the protocols were followed.[15] Thus, the intervention provided treatment firms with the opportunity to produce rugs for the export market.

We allowed Hamis to allocate post-treatment orders for two reasons. First, it was infeasible for us to demand that Hamis continue to work with a firm that was clearly not able to produce at an acceptable standard. Hamis' foreign buyers are demanding and would not accept subpar rugs. Second, for external validity purposes, we wanted the experiment to mimic a normal buyer-seller relationship as closely as possible. Our intervention places initial orders with a random set of producers, but allows the intermediary to optimally allocate further orders *within the treatment group* based on firm quality, reliability and so forth. As such, subsequent orders are endogenous. However, whether a firm is in the treatment group and hence offered the opportunity to export, is, of course, random and allows us to identify causal impacts of exporting (via the 'intent-to-treat' specification we will discuss below).

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. From a theoretical point of view, trade models typically model exporting as a demand shock, sometimes with features distinct from domestic demand shocks. Increasing demand is also the primary motivation for many export facilitation policies (e.g., sending trade delegations, researching foreign markets, building export infrastructure such as ports or streamlining export regulations). Therefore, to assess the impacts of exporting, it is natural to include this central component. In terms of the practical limitations, if we were to provide equally-sized domestic orders it is unclear on what dimension they should be equal given the different profit margins and hours required per rug. Even then, it would have been extremely difficult to acquire anything like the $155,682 of firm orders (and a larger number still at the prices

[15] 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.

the intermediary received) that came from international markets.

3.2 Experiment Takeup

The third row of Table 1 shows the takeup status for Sample 1 (columns 1-4). As anticipated by our decision to stratify along this dimension, takeup rates varied greatly depending on the firm's primary rug type. For goublain and tups producers, the two rug types for which we obtained no orders, take-up rates are 10 and 19 percent, respectively. We expected low takeup values in these strata since these firms do not typically produce duble or kasaees rugs. Nevertheless, we attempted to treat these firms and found that very few were willing to switch rug types.[16]

In contrast we did have export orders for kasaees and duble rugs. Table 1 shows that among kasaees and duble rug producers take up was 26 and 38 percent, respectively, but the takeup rates were still relatively low. As previously mentioned, and shown in Figure 4, between December 2010 and May 2012, ATA and Hamis were unable to secure a large number of export orders even for duble rugs. As a result, we were unable to approach treatment firms in Sample 1 with the opportunity to produce 110m^2 in one go. Instead, we had to offer smaller orders of 20m^2 sequentially, or about two weeks of work. Because this initial order size was small, many firms were unwilling to work with us.[17]

From March 2013, Hamis' major buyers offered assurances that they would continue to place duble rug orders for the foreseeable future so it was possible to offer the opportunity to produce 110m^2 in one go. We therefore decided to draw a second sample of firms that just produced duble. Given that all our export orders were for duble rugs, this would increase the sample size of duble firms substantially. Additionally, given the larger order size, we expected higher takeup within duble producers. In February 2013 the survey team found an additional 140 firms that specialized in duble production and were not in the original listing exercise; we refer to these firms as "Sample 2". As with the initial sample, we stratified these firms on loom size and 35 firms out of the 140 were randomized into the treatment sample.[18]

Given the large export orders that Hamis had secured, we could now offer the full 110m^2 at once to the treatment firms in Sample 2 and we could also ensure that all the treatment firms in Sample 1 received their full 110m^2 allotment. As anticipated, this large order led to substantially higher takeup in Sample 2 firms. Column 5 of Table 1 reports treatment and takeup statistics for Sample 2: 32 out of 35 treatment firms agreed to produce the export orders for Hamis.[19]

The 5th row of Table 1 reports the number of "successful" takeup firms, defined as those who

[16]During the survey round 2, we asked firms why they did not takeup. Appendix Table B.1 confirms that the main reason for refusals among goublain and tups firms was that the export rug order was not the suitable rug type.

[17]As shown in Appendix Table B.1, many duble firms report being unwilling to jeopardize their existing relationships with other intermediaries for a small amount of work. Other duble treatment firms reported that the export order was not the suitable rug type as they misreported duble as their primary rug type at baseline. Many kasaees producers were unwilling to accept the export order because the particular rug they were asked to produce was different from the kasaees rugs they usually make.

[18]The choice of 35 treatment firms for Sample 2 was dictated by Hamis' constraints on the number of firms it could work with, and our desire to ensure that the full 110m2 could be offered to each treatment firm.

[19]30 of the 35 Sample 2 treatment firms took the offer up immediately in March 2013. The 2 remaining firms began producing orders for Hamis in May 2014. This delay was due to capacity constraints on the side of Hamis.

produced more than 110m^2 and received subsequent orders from Hamis. As shown in Table 1, only 4 treatment firms (all in Sample 1) failed to secure additional orders from Hamis after the initial treatment. Two of the firms were unable to manufacture the export orders successfully while the remaining two firms had a falling out with the owner of Hamis. The fact that the overwhelming majority of firms were successful is itself interesting, and is likely related to the learning-by-exporting results we document below.

Given that we only generated large and sustained export orders in one rug type, duble, and that very few firms outside of duble were willing to manufacture this rug type, we do not include the non-duble strata in our analysis.[20] To be clear, if the focus of this paper was to simply evaluate the trade facilitation program, it would be important to understand why the intervention only generated sustained exports for one of the four products (and to answer that question, our randomization would have to be over many products not many firms). Instead, our paper asks a different question that is central to the learning-by-exporting literature: does exporting improve firm productivity?

In terms of the analysis, Sample 2 has two advantages over the duble firms in Sample 1. First, as noted above, Hamis was able to offer large initial treatment orders all at once to Sample 2, which resulted in much higher takeup rates. This means that there was less potential selection among takeup firms (recall 32/35 treatment firms took the offer) which affects the interpretation of the treatment-on-the-treated specifications. Second, the treatment in Sample 2 is the treatment we intended when designing the experiment. Firms in Sample 2 were offered a large initial order followed up by continued orders if the initial order was filled satisfactorily; in contrast Sample 1 firms did not receive a reliable flow of orders until 1.5 years after the beginning of the study. This fact can be seen from Figure 4 which superimposes the dates of the survey rounds on Hamis' cumulative exports. For these reasons, we will present two sets of results, the first restricting the analysis to Sample 2 firms only (with 140 duble producers, 35 in treatment and 32 who took up), and the second pooling all the duble producers in Sample 1 and Sample 2 (the "Joint Sample" with 219 duble producers, 74 in treatment and 47 who took up). Since Sample 2 is our preferred sample, we focus our discussion on the results for Sample 2 and note any discrepancies with the Joint Sample when they arise.

3.3 Data

Data collection for each sample occurred in three phases: baseline, periodic follow-up surveys and endline. In both the baseline and endline we collected data on: (a) firm production; (b) rug quality; and (c) household and demographic characteristics. All nominal variables are converted to real values using the Egyptian CPI. In the follow-up surveys we only collected data on firm production and rug quality. The initial intention was that follow-ups surveys would be conducted quarterly but political turmoil in Egypt resulted in several unanticipated delays.[21] Table 2 shows

[20] Although we did initially obtain some kasaees export orders, we did not manage to obtain sustained orders. Given our inability to generate sustained orders we also ignore these strata in the analysis.

[21] On three separate occasions, political turmoil meant that Egypt's Central Agency for Public Mobilization and Statistics stopped processing our applications to continue surveying in Fowa.

the timeline of surveys for both samples.[22]

The production module records production activity for the month preceding the survey interview. 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: (1) the type of rug being produced; (2) how difficult the rug is to make rated on a 1-5 scale by a master artisan (see below); (3) the amount of thread used per m^2 of the rug (thread count); (4) the number of colors used in the rug; and (5) which segment of the market the rug is aimed at as reported by a 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:[23] (1) corners; (2) waviness; (3) weight; (4) touch; (5) packedness; (6) warp thread tightness; (7) firmness; (8) design accuracy; (9) warp thread packedness; (10) inputs; and (11) loom.[24] Each measure is rated 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, a flatter-lying rug or a more accurate design are attributes valued by both foreign and domestic consumers. As discussed in Section 2.2, higher quality scores along most dimensions reflect higher weaving skill and technique.

For takeup firms, a second quality module is available at higher frequency. These firms deliver rugs to Hamis on a weekly basis. Upon receiving the rugs, Hamis checks the rugs for size accuracy, design accuracy, packedness, firmness, weight and records how "ready" the rug is for final delivery. Less ready rugs require various efforts by the intermediary to improve the look and feel, such as cutting off loose threads or fixing threads to reduce the waviness of the rug. High-quality rugs do not require much time to ready for delivery, hence we interpret this measure as an indicator of quality.

We collected a third set of quality measures in June 2014 by setting up a quality lab in a rented workshop where firm owners were brought and asked to produce an identical domestic-specification rug using identical inputs and a common loom. The rugs were anonymized and scored along the quality dimensions listed above by both the master artisan and a Professor of Handicraft Science from Domietta University located 2 hours east of Fowa.

Finally, we administered a household module at baseline and endline. This module collects information on household income, literacy rates and so forth.

[22] We hired an Egyptian survey company to conduct the baseline survey on Sample 1. The company also trained an enumerator who was responsible for follow-up Round 1. Unfortunately, we discovered that this enumerator made up much of the Round 1 data, and so this round has been discarded. We immediately fired the enumerator and hired new employees in January 2012 and conducted all subsequent surveys.

[23] The first Sample 1 survey round recorded 6 quality metrics to which we subsequently added 5 more metrics.

[24] Corners captures the straightness of the rug edges. 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 can have 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.

3.4 Summary Statistics

Table 3 shows baseline balance between the treatment and control groups for Sample 2 in the left panel and the Joint Sample in the right panel. 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 treatment and control is around 50 years and, on average, firms have slightly more than 35 years of experience working in the rug industry. Roughly 60 percent of firm owners are illiterate. The average household size is 4.

Panel B reports statistics from the rug business. Monthly profits from the rug business averages LE874 ($134 using the exchange rate of 6.51). Firms report 268 labor hours in the previous month, which amounts to around 22 days of work at 12 hours per 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 43.5m^2 and only about 16 percent 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 and samples we find no statistical differences between treatment and control firms with one exception: in the Joint Sample, treatment firms report *lower* quality scores at baseline. The final row of Table 3 reports attrition across survey rounds. Attrition has been low with a non-response rate of approximately 4 percent per round (11 percent for the Joint Sample) which does not vary across treatment and control groups .

4 Causal Impacts of Export-Market Access on Profits

4.1 Empirical Specifications

The randomization methodology allows us to adopt 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. Since (1) controls for the baseline value of the dependent variable, we cannot include observations from the baseline survey in the regression.[25] Since, as discussed in Section 3, not all firms who were offered the opportunity to export took up that offer, (1) is an intent-to-treat (ITT) specification.

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

[25] Alternatively we could use all survey rounds, include firm fixed effects, and interact $Treatment_i$ with a post-baseline dummy. We chose a specification with baseline controls because if the dependent variable is measured with noise and we have relatively few survey rounds, the fixed effects estimator will perform poorly (see, for example, Karlan and Valdivia (2011) and Angelucci et al. (2015)).

the treatment effect to take account of the fact that not everyone was actually treated:[26]

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

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, in so far 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 4, being in treatment raises the probability of ever exporting by 68 percentage points from a baseline of 19 percent in Sample 2 and 55 percentage points from a baseline of 13 percent in the Joint Sample. We also report the TOT specification, which suggests even more dramatic increases.[27]

4.2 Measuring Profits

Profits are notoriously difficult to measure, particularly for firms who do not keep regular accounts. As a result, de Mel et al. (2009) use several methods to elicit profit measures from small firms. Their analysis suggests that there is often a mismatch of revenues with the expenses incurred to produce those revenues; for example, if there are lags between incurred material expenses and sales, asking revenues and expenses in a given month will not accurately capture firm profits. They advocate asking firms to directly report profits instead.

Following de Mel et al. (2009), we construct four measures of profits. The first measure is a direct profit measure from the firm's response to the question: "What was the total income from the rug business last month after paying all expenses (inputs, wages to weavers but excluding yourself). That is, what were your profits from this business last month?" The second measure constructs profits from two surveys 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 idea behind this measure is that there may be less noise in constructing profits from its components–prices and quantities–than from recall information on total revenues and expenses; we refer to this measure as "constructed profits". This measure is also free of the concern that firms might use business expenses for household consumption (or use business revenues to pay for household expenses) that may be an issue for the other two measures. Finally, we construct a fourth measure based on a hypothetical question that asks firms how much they would earn from selling a specific quantity of inputs. Specifically, we construct "hypothetical profit" by asking firms how much it would cost

[26]For Sample 2, the ITT and TOT will be very similar given the high takeup. For the Joint Sample, the TOT will be an upper bound if the firms who took up the intervention were the ones with most to gain from exporting.

[27]Note that the ITT and TOT do not scale up by the takeup rates shown in Table 1 since a handful of firms that eventually took up had not done so yet at the time of some early survey rounds.

to purchase 25 kilograms of the thread they used in the previous month, how long it would take to weave this output, and how much they would earn from selling the output. Although not the realized profits of the firm, this measures alleviates potential concerns regarding the timing of when revenues are earned and costs are incurred and serves as a check against the three profit measures.

4.3 Profit Results

Table 5 shows the results of running specifications (1) and (2) on logged values of the four profit metrics. We first discuss the top two panels which report results using Sample 2 and the Joint Sample, respectively. As before, we discuss the results from our preferred sample, Sample 2, and note if there meaningful differences in the Joint Sample. The columns display different profit measures as outcome variables and for each we report the ITT and TOT.

The first two columns (1A and 1B) report the specifications using the (log) direct profit measure. The ITT coefficient is 0.25, implying that the export treatment increases monthly profits by approximately 25 percent. The TOT coefficient is, not surprisingly, larger at 30 percent and also statistically significant. The Joint Sample has a similar ITT but a larger TOT estimate of 42 percent.

Columns 2A and 2B report specifications using a profit measure constructed from asking firms about total revenue and costs in the previous month. We observe very similar point estimates: the ITT and TOT are 23 and 28 percent, respectively. The reason these point estimates are similar to columns 1A and 1B may be because the firms in our sample typically do not store much inventory (hence the timing mismatch between revenues and expenses is not severe) and rug inputs are unlikely to be used for household consumption.

We report the results using constructed profits in the columns 3A and 3B. The point estimates are again very similar to the previous columns: the opportunity to export raises profits by 24 percent. Finally, we examine the "hypothetical profit" measure in columns 4A and 4B. These estimates are higher than the previous numbers. The ITT point estimate is 36 percent. It is reassuring to see consistency across all four measures. We also note that the treatment effects reflect profits rising among treatment firms rather than falling profits among control firms; profits for control firms remained flat in real terms across pre- and post-baseline survey rounds.[28]

These regressions indicate that the export treatment causally increases profits by between 23-36 percent. 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 do not report a wage paid to the owner. Therefore, in Panels C and D of Table 5, we construct profits per owner hour by dividing each profit variable by the total hours worked by the owner (or other unpaid family members) in the previous month. Using the direct profit per owner hour measure in columns 1A and 1B, we find that the ITT estimate is 17 percent. 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. The differences between Panels A and C suggest

[28]Regressing profits on a post-baseline dummy yields a coefficient of 0.049 (s.e. 0.049) when we restrict attention to Sample 2 control firms only (0.041 (s.e. of 0.055) for the Joint Sample).

that total owner hours increased by 6 to 8 percent, depending on the sample.[29] In the subsequent section, we examine this increase in hours in more detail. Nevertheless, the basic message remains the same: the opportunity to export raised profits per owner hour by 15-21 percent.

Before turning to mechanisms, and in particular whether or not these improvements in firm performance occur through learning-by-exporting or another mechanism, 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. Business training had a statistically significant impact on profits in only two out of nine studies that measured profits. One possible interpretation of the mixed results is that investments in management and production practices may only be effective in the absence of demand constraints. For example, the returns to business literacy may be low if there is insufficient demand. Our results suggest a potentially important role for relaxing demand constraints through expanding market access. Another popular intervention normally targeted at small firms is expanding access to credit. The literature on the impacts of credit on profits for small firms also finds mixed results. For instance, de Mel et al. (2008) find returns to capital of around 5 percent per month while Banerjee (2013) cites several credit interventions that produced no statistical increases in profits. As such, our evidence suggests that demand constraints may be an important factor limiting the growth of small firms.

5 Sources of Profit Changes

5.1 Prices, Output, Input Factors and Costs

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

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

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 and F is a fixed cost of production. We do not include input costs in (3) since 96 percent of Sample 2 firms (and 74 percent of firms in the Joint Sample) receive raw material inputs from their intermediary and hence do not pay for these expenses. Hamis follows this industry norm. For the small percentage of firms that do purchase inputs on their own account, we subtract the prices of the warp and weft thread inputs from p to make these prices comparable across all firms.

Table 6 uses our survey data to examine the various components of the profit increase shown in the previous section. Columns 1A and 1B evaluate the impact of the intervention on the log output price. The ITT specification indicates 46 percent increase in prices with the opportunity to export while the TOT indicates a 56 percent increase. Thus, part of the profit increase from exporting is coming from significantly higher prices per m^2 of rug for export orders.

[29] For the hypothetical measure in columns 4A and 4B, we divide hypothetical profits by a hypothetical measure of how long the firm would take to weave 25 kilograms of thread. This is why the difference between columns 4A and 4B in Panels A and C (or B and D) does not match the increase in total hours inferred from the other columns.

Columns 2A and 2B 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 -22 percent while the TOT is -27 percent; this is a large *decline* in output among treatment relative to control firms.[30]

Columns 3A and 3B document the impact of the intervention on firm scale, as captured by the log of total hours l worked by all employees in the firm in the previous month. The ITT estimate indicates an increase of 8 percent and the TOT is 10 percent. 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). The fact that expansion occurs primarily along the intensive margin suggests there may be large discontinuities in the cost of hiring additional workers, particularly since an additional weaver is likely to require his own loom.

Finally, we turn to fixed costs F in columns 4A and 4B. The main proxy we use is the size of the warp thread ball, measured in (log) kilograms, that is placed on the loom at the beginning of a production run. A larger warp thread ball enables firms to amortize the costs of re-stringing the loom over more units. The ITT estimate is 13 percent indicating that the opportunity to export lowers the fixed cost of a production run by running longer runs that require less frequent re-stringings of the loom.

Table 7 examines input prices and quantities. As noted above, 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 explore the impact of the intervention on reported weft and warp thread prices. Recall that the weft thread is used to create the pattern of the rug and the warp thread is the base thread. Reported weft thread prices increase 23 percent. 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 material type (cotton, wool, polyester, silk or various blends), thickness and material grade (e.g., Egyptian wool or more expensive New Zealand wool). Note that although columns 3-4 suggest that input quantities (measured in grams) do not increase with the opportunity to export, the output decline implies that exported rugs use more material inputs and are heavier than domestic rugs.

5.2 Interpreting the Sources of Profit Changes

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

[30]The sum of the point estimates on prices and output does not precisely match column 3 in the previous table because of differences in sample sizes due to missing observations..

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 prices (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. The rise in material input prices provide further evidence for such an explanation if high-quality rugs require more expensive high-quality inputs (Kugler and Verhoogen 2012). In the next subsection we confirm this conjecture.

5.3 Quality and Productivity

We first draw on the detailed quality metrics described in Section 3.3 to confirm that treatment firms are indeed manufacturing higher quality products. We have 11 different quality metrics that are ranked on a 1-5 basis with 5 being the best for that type of quality.

Table 8 presents the results for the quality metrics. Instead of implementing specification (1) or (2) separately for each quality metric, we regress a stack of all 11 quality metrics on interactions of the treatment (or takeup, for the TOT) with indicators for each of the quality metrics. We also include interactions of the quality-metric indicators with baseline values, quality metric fixed effects, as well as both the strata and round fixed effects. The coefficients from this regression are identical to running separate regressions for each quality metric, but allows us to cluster the standard errors by firm to account for possible firm-level correlations either within a quality metric across time or across quality metrics within a time period.

For 10 of the 11 quality metrics, quality is significantly higher among treatment firms. The one exception is the quality of the loom, where we find no treatment effect. 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 8 restricts the coefficients dummy to be identical across the various quality metrics (recall they were all run in a single stacked regression).[31] 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 1.14 points higher among treatment firms. These are substantial increases in quality given a standard deviation of quality of 0.55 at baseline.

We also examine productivity, measured as *output per labor hour.* This measure comes from firms' responses to the question: "how long does it take you to make 1 meter squared?".[32] This

[31] This method is similar to estimating the impact of treatment on a standardized index of quality measures in each survey round (e.g. Kling et al. 2007), but we prefer our method as it produces more conservative estimates in our data (i.e., higher standard errors).

[32] Another way to measure output per hour is to divide total output by total hours worked in the month. We believe the direct measure is less noisy and so we use it for the analysis. We find virtually identical results using the alternative measure (available on request).

productivity measure is based on the production technology described in Section 2.2. The production technology is Leontief in labor and materials. Labor is the primary input and materials are non-binding since the majority of firms are provided inputs by their dealers. We abstract from capital because there is very little variation in capital across firms: 92 percent of firms use one loom (and 98 percent in Sample 2) and no firm in our sample purchased (or rented) an additional loom since the beginning of the study.[33] We also consider a second productivity measure that relaxes the assumption that only labor is required for production. We estimate total factor productivity (TFP) using a Cobb-Douglas production function with both labor and capital, and accounting for the simultaneity of input choices (see Appendix A for further details).

Panel A of Table 9 shows the results for output per hour. The ITT estimate indicates that output per hour falls 24 percent among treatment relative control, with even larger TOT effects. Panel B presents the TFP measure and we find a similar 29 percent decline for the ITT specification.

5.4 Mechanisms

The finding that quality rises and output per hour falls alongside rising profits is consistent with two different 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. While it is quite challenging to provide a direct mapping between profit margins and quality levels, we provide some suggestive evidence for this phenomenon by analyzing Hamis' (self-reported) cost structure for domestic and foreign orders. Hamis reports 9 percent profit margins on domestic orders and substantially higher margins of 33 percent on foreign orders with the full cost structure broken down in Appendix Table B.2. This provides some evidence that the higher prices we observe among treatment firms may come from these profits being shared between Hamis and the producer. Under this mechanism, the export opportunity raises the relative price of high-quality rugs and profit-maximizing firms respond by producing rugs to specifications associated with high-quality. What does *not* change through this mechanism is technical efficiency.

A second mechanism is learning-by-exporting, which we follow the literature and define as an export-induced change in technical efficiency (Clerides et al. 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 towards high-quality production, quality upgrading can also occur through these learning processes.

We emphasize that the two channels *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 increases in profits beyond those generated by simply moving along the PPF. In the

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

next section, we define learning-by-exporting in a more precise manner and provide evidence it is present in our setting.

6 Detecting Learning-by-Exporting

In order to be explicit about the learning-by-exporting mechanism, we enrich the profit function by detailing production functions for output and quality:

$$\max_{l,m,\lambda} \pi = p\left(q(\lambda)\right) x(\lambda, l) - wl - F \tag{4}$$

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

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

$$p = p_0 + bq \tag{7}$$

where p is now a price function that is exogenous to the firm and depends on a quality-unadjusted component p_0 and on the quality of the rug q, with $b > 0$.[34] Rug output and quality are determined by two production functions, both of which depend on a choice variable: the product specifications of the rug indexed by λ. (Recall that specifications are codifiable rug characteristics that buyers agree upon before ordering; see Figure 2 for an example of such an agreement.) High-λ rugs have more demanding specifications, in the sense that they require more labor hours to produce, and we assume that these high-λ specifications are also associated with high-quality rugs.

The production function for output $x(\lambda, l)$ has two components. Labor inputs are mapped to output through $f(l)$ and output per unit of labor input is determined by the function $a(\lambda; \chi_a)$, an output productivity measure that is "unadjusted" for rug specifications.[35]

Output productivity $a(.)$ is necessarily decreasing in λ since rugs with more demanding specifications require more labor hours. The function $a(.)$ is also increasing in χ_a, an output efficiency parameter. 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. Additionally, quality increases in a quality 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 shift along the PPF and what constitutes a shift out (i.e. learning-by-exporting). Firms shift along the PPF when there is an increase in b, the price of quality. This leads firms to choose higher specifications λ, and by (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

[34] See Verhoogen (2008) for a microfoundation for this relationship between price and quality.

[35] We choose this simple parametrization since, as discussed in Section 5.3, there is little variation in capital across firms and the production technology is Leontief in materials and labor, with labor the binding factor.

occur either as firms move into high-quality products with steep learning curves and/or through transfers of knowledge between foreign buyers and domestic sellers. We might expect transfers of knowledge about quality, χ_q, to be particularly relevant for firms in low-income countries that sell to buyers in high-income countries that have have more sophisticated tastes and demand higher-quality products. Despite the different theoretical implications, we are unaware of earlier work that seeks to distinguish these two mechanisms of quality upgrading.

To see that this theoretical framework can generate reductions in output per hour alongside quality improvements through either of these mechanisms, consider the following functional forms for the production functions:

$$a(\lambda;\chi_a) = (\chi_a - \lambda)^{\alpha} \qquad \alpha \leq 1 \tag{10}$$

$$f(l) = l^{\beta} \qquad \beta \leq 1 \tag{11}$$

$$q(\lambda;\chi_q) = \lambda\chi_q \tag{12}$$

where λ is complementary with χ_a and χ_q in producing output and quality respectively (weakly so in the former case). Maximizing (4) using the functional forms in (10)-(12) yields the equilibrium product specifications and labor inputs, and hence equilibrium quality and output productivity:

$$q^* = (\frac{\alpha}{1+\alpha})\left(\frac{\chi_a\chi_q}{\alpha} - \frac{p_0}{b}\right) \tag{13}$$

$$a^* = (\frac{\alpha}{1+\alpha})^{\alpha}\left(\chi_a + \frac{p_0}{b\chi_q}\right)^{\alpha} \tag{14}$$

By taking derivatives of the equilibrium values with respect to either the price of quality (b) or the efficiency parameters (χ's) it is easy to see that both mechanisms can generate the results we found in the previous section. Increases in b or the quality efficiency parameter, χ_q, lead firms to raise product specifications and hence produce higher quality products that sell for higher prices. However, output and output productivity will fall alongside rising labor demand because high-λ rugs require more labor inputs. As discussed in Section 5.4, a rise in b may be a precondition for learning-by-exporting to occur as buyer knowledge may be specific to, or the learning curves steeper for, the high-quality rugs foreigners demand (as captured by the complementarity between λ and the χ parameters).

For clarity, our model does not allow for investments that raise the χ parameters. If the return to these investments rises with the opportunity to export, any resulting changes in the χ parameters would not be considered a shift out in the PPF and hence should not be classified as learning-by-exporting under our definition. 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 not paid for by the firm, even implicitly, would be. 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 very difficult to disentangle treatment effects of exporting from selection (Melitz, 2003). The most convincing analyses to date rely on matching techniques which requires an assumption 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 total factor productivity. If prices are higher in export markets, productivity 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.[36] In the few cases where price adjustments are made, measuring quantity-based productivity requires comparing products with identical specifications. This is typically achieved by relying on product dummies (e.g., de Loecker et al., 2014), with the extent of disaggregation determined by administrative classifications, or focusing on homogenous goods like concrete and block ice (e.g, Foster et al., 2008). In contrast, we exploit our rich panel data and experimental variation to solve these issues.

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

1. In Step 1, we show that although output productivity falls with the opportunity to export, productivity *conditional* on rug specifications *rises*. We also show that quality levels rise conditional on rug specifications. If there is no learning-by-exporting, output productivity and quality levels should be *unchanged* once we condition on λ.

2. In Step 2, we demonstrate that when asked to produce *identical* 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 not differ in rug quality when producing identical domestic rugs.

3. In Step 3, we use time-series data to establish that quality and output 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 output productivity levels.

4. In 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 quality most along the particular quality dimensions that are discussed during meetings with Hamis. This evidence also allows us to show that learning-by-exporting is not driven by learning-by-doing (triggered by the export orders) alone, but in part through transfers of knowledge.

5. In Step 5, we rule out the alternative hypotheses that there were adjustment costs or that firms made investments to raise their efficiency parameters. In particular, we show that treatment firms make no monetary or time investments in upgrading; do not pay, even implicitly, for the knowledge they receive from the intermediary; and the costs of providing

[36] See de Loecker (2011) for an extensive discussion of this point.

such knowledge are far less than the increased profits it would generate even in the domestic market. This evidence suggests that either firms did not know this knowledge was obtainable or there are failures in the market for this knowledge.

6.1 Step 1: Conditioning on Rug Specifications

If firms are only moving along the PPF, changes to quality levels and productivity should occur only through changes in rug specifications: $\frac{da}{db}|_{\lambda}, \frac{dq}{db}|_{\lambda} = 0$. That is, producers know precisely 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: $\frac{da}{d\chi_a}|_{\lambda}, \frac{dq}{d\chi_q}|_{\lambda} > 0$.

Thus, to detect learning-by-exporting, we repeat the quality and productivity regressions above but now control for the specifications of the rug being manufactured at the time of the survey visit.[37] Although we are not able to control perfectly for the myriad of possible rug specifications, we include the five specifications described in Section 3.3 as controls. We note that many studies 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 ten-digit classification 580500200 (hand-loomed wool rugs)). Moreover, we include additional specifications, such as thread count or design difficulty, that are rarely observed by researchers. For example, we know the if the rug is destined for low-tier or high-end stores. Including these controls is possible because there is overlap in rug specifications across firms selling to domestic and foreign markets. This overlap can be clearly seen in Figure 5 which plots the distribution of each of the five specifications both for firms that are producing rugs for export (i.e. $Takeup_{it} = 1$) and for those that are not. Note that if our characteristic controls are very crude, that will tend to bias our findings towards the unconditional results we found in Tables 8 and 9. Hence, the prediction that output productivity should rise conditional on specifications is particularly informative since output productivity fell in the absence of controls.

We first explore the stacked quality measures in Panel A of Table 10. Focusing on Sample 2 in columns 1-2, we see that the ITT and TOT estimates are positive and statistically significant. Reassuringly, the specification controls have the signs we assumed in the model. More difficult rugs are associated with higher quality while those destined for lower segments of the market are associated with lower quality. We present the results for each quality metric in Appendix Table B.3 (left panel) and the conclusions are unchanged.

The bottom panel of Table 10 reports the results for productivity. Recall that Panel A of Table 9 reports an ITT estimate of *negative* 24 percent, and this changes to *positive* 28 percent when we condition on rug specifications in Panel B of Table 10. That is, conditional on the five rug specifications, the opportunity to export is now associated with significantly higher output productivity.

[37] Since all our regressions also include controls for the baseline values of the dependent variable, we must also include baseline values of the specifications in the controls when running specifications with characteristic controls.

The adjusted R-squared triples (from 0.18 to 0.53 in the ITT estimates) suggesting that the rug specifications have substantial explanatory power. The estimates using TFP in Panel C exhibit a similar reversal. If firms are just moving along the PPF, there should be no changes in output productivity conditional on product specifications. The data strongly suggest otherwise. Once again, the rug characteristic controls have the signs we assumed in the model: rugs with more colors and those that require more thread per m^2 require more labor inputs. Relative to products at the high-end market segment, lower segment rugs require fewer labor inputs.

A reasonable concern with this exercise is that, although treatment is exogenous by design, the specifications controls on the right-hand side may be endogenous. An alternative approach is to adjust the productivity and quality measures for specifications, and regress the specification-adjusted measures on treatment. To perform this adjustment, we first regress productivity (or quality) on the rug specifications at baseline, before any experimental intervention, and use the resulting coefficients to construct adjusted productivity (actual minus predicted productivity) for each round.[38] Of course, if higher-ability firms selected into higher-specification rugs at baseline, the coefficients on rug specifications will be biased. However, if anything this bias will lead us to find no productivity gain.[39] Table 11 shows the ITT results from regressing the specification-adjusted quality and productivity measures on treatment (or on takeup instrumented with treatment for the TOT).[40] Reassuringly, we find that adjusted productivity and quality rise by similar magnitudes as before, suggesting that the endogeneity of specifications is not driving our results.

6.2 Step 2: Production of Identical Domestic Rugs (the Quality Lab)

The second step exploits our experimental setting to compare quality and productivity across firms producing identical domestic rugs (rather than relying on specification controls to control for the type of rug). If firms are only moving along the PPF, when asked to make identical domestic rugs, quality and productivity should not differ across treatment and control firms (since treatment was randomly assigned). In order 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 domestic-specification rug using identical inputs and the same loom.[41] We chose rug specifications that mimicked a popular rug design sold at mid-tier domestic retail outlets in Egypt. The rug was to be 140cm by 70cm with a desired weight of 1750g, and the master artisan assigned a difficulty rating of 3 for this rug (below the 4.28 average rating of export orders).

After all firms had manufactured the rug, each rug was given an anonymous identification number and the master artisan was asked to score each rug along the same quality dimensions discussed previously. The identification system ensured that the master artisan had no way of

[38]For the baseline of Sample 1, we did not record the market segment or the rug difficulty. We replace these missing values with the corresponding values from the subsequent survey round.

[39]Specifically, due to selection the productivity penalty for making a high-specification rug may be larger than the coefficients imply. If our experiment induced lower-ability firms to make high-specification rugs, the ITT comparing specification-adjusted productivity between treatment and control may be biased downward (because lower-ability firms in the treatment group would "appear" less productive if the coefficients used for the adjustment were biased).

[40]The right panel of Appendix Table B.3 reports the estimates separately for each adjusted quality metric.

[41]The owner was paid a market wage plus compensation for having to make the rug in an external location.

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, to cross check the accuracy of the master artisan's scoring.

In Panel A of Table 12, we report results separately for each quality metric.[42] Quality is significantly higher among treatment firms for all 9 quality dimensions.[43] Reassuringly, treatment firms also score higher along every dimension using the professor's quality assessments.

Panel B of Table 12 reports the results constraining the coefficients on treatment to be the same across all the quality metrics. Given the results from Panel A, it is not surprising that the coefficients are statistically significant. The point estimate from column 1A of Panel B is 0.86. Given that the standard deviation of the master artisan's quality metrics is 0.84, the point estimate suggests that treatment firms produce at quality levels a full standard deviation above control firms.

Panel C of Table 12 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. We do not observe statistical differences in the width of the rugs, but this is expected since the loom size determines the width (and all firms used the same loom). The third row shows that treatment firms also produce rugs that are closer to the requested weight.

To measure productivity, we recorded the time taken to produce the rug. Again, since the rug and loom are identical across firms in this setup, the time taken reflects firm productivity. The 4th row of Panel C shows that, on average, firms took 4 hours to produce the rug and there is *no* difference in the time taken across treatment and control firms. That is, despite manufacturing rugs with higher quality metrics, treatment firms did not spend more time weaving: quality-adjusted productivity is higher.

In the absence of learning-by-exporting, we would not expect differences between treatment and control firms when producing identical 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 quicker or at higher quality since they are more used to manufacturing domestic designs and specifications. 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. Moreover, treatment firms do not take longer to produce these rugs. 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: It is just as plausible that control firms put in extra effort to impress Hamis in order to gain export orders, and if treatment firms were putting in more effort, we would expect them to take longer to manufacture the rugs. Thus, the finding that productivity is the same, but not higher, is consistent with the conjectured complementarity between specifications and the output efficiency parameter χ_a (so that productivity gains are

[42] As before, we account for correlations across quality metrics by stacking our metrics, interacting treatment with each metric and strata fixed effect, and clustering standard errors by firm; this alters the standard errors but not the magnitudes of the coefficients.

[43] We have 9 quality metrics since loom quality and input quality are not relevant in this setting because all firms used the same loom and were provided the same inputs.

concentrated in initially less-familiar export rugs).

6.3 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 6.5 for a discussion of adjustment costs), learning processes typically take time. Hence, we write the two efficiency parameters in period t as:

$$\chi_{k,t} = h_k \left(\sum_{t'=0}^{t} \left(x_{t'} \mathbf{1} \left[\text{export}_{t'} \right] \right) \right) \text{ for } k = q, a. \tag{15}$$

where $\mathbf{1}\left[\text{export}_t\right]$ is a dummy that takes the value of one if the rug output that period is for export. In this 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 the firms. Therefore, if there is learning-by-exporting, productivity and quality should rise with cumulative exports. If there is no learning, $\chi_{k,t} = \chi_{k,0}\ \forall t$, although quality may immediately jump (or unadjusted productivity may fall) with the first export order, the levels should remain constant with additional export orders.

To investigate potential learning curves in a non-parametric manner, we carry out a two-stage procedure. In the first stage, we regress either quality or productivity on firm fixed effects as well as round fixed effects.[44] 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 just include these firms in the second stage (although all firms are included in the first stage when we de-mean by survey round). As previously, we focus on Sample 2 firms.[45] We also show similar plots that either include product specifications in the conditioning set in the first stage, or that use the specification-adjusted measures described in Section 6.1 as the dependent variable in the first stage. Finally, Appendix Figure B.2 presents similar plots using the partially linear panel data estimator proposed by Baltagi and Li (2002).

Figure 6 shows these residual plots for both productivity measures as well as the stacked quality measure. The upper left graph reports the output per hour measure; the figure indicates a decline in output productivity until about 400m^2 after which output per hour starts to rise. We draw similar conclusions from the TFP measure (middle-left figure). The rightmost panels include product-specification controls or use specification-adjusted productivity. In these cases, productivity rises with cumulative exports, consistent with the initial dip in unconditional productivity being driven by the move to more difficult product specifications demanded by foreign buyers.[46]

The bottom row of Figure 6 presents the analogous learning curve for the stacked quality mea-

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

[45] Analyzing the Joint Sample here is difficult. First, Sample 1 firms had much longer gaps between orders and smaller order sizes in the first year of the project and so cumulative exports is a far more noisy measure of the stock of knowledge. Second, given that Sample 1 firms started exporting earlier they drive all the variation at high values of cumulative exports but not at lower values.

[46] The one exception is the TFP measure using specification controls where there is an insignificant and moderate downward slope up to 400m^2.

sures. The patterns show a sharp rise in quality by 200m^2 of exports and then levels off. The typical firm weaves about 10-15m^2 per week, which suggests that firms learn how to produce the quality demanded by foreigners within about three months. In the right two panels, we observe a similar path when including product-specification controls or using specification-adjusted quality, indicating that the pattern is not driven by changes in product specifications alone. The figures suggest much faster learning about quality efficiency $\chi_{q,t}$ than about output efficiency $\chi_{a,t}$.[47]

The quality measures used in Figure 6 were recorded by the master artisan at the time of each survey. For the subset of firms that produced orders for the intermediary, we have an additional set of quality metrics for each batch of rugs delivered by each firm (often at a weekly frequency). We can produce similar plots for the 6 high-frequency quality measures recorded in this manner. Figure 7 shows local polynomial regressions of the residuals for these metrics (after regressing each measure on firm fixed effects as before) with Appendix Figure B.3 presenting similar plots using the partially linear panel data estimator. For 4 of the 6 metrics (size accuracy, firmness, packedness, design accuracy) we observe similarly quick learning curves as in the master artisan data. For the remaining two metrics, the readiness of the rug for delivery and weight accuracy, we observe more limited evidence of learning.[48]

6.4 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 is driven by knowledge transfers.

The control we have over our experiment allows us to record and measure knowledge flows. We observe information being transferred between both 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 and are more suggestive in nature. Here we provide several excerpts documenting information flows between overseas buyers and Hamis regarding specific 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:

[47] 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 et al. (2013) document a 70 percent decline in defect rates in an automobile manufacturing firm just 8 weeks after new production processes were introduced.

[48] For rug readiness, 5 is the most ready and 1 the least. Weight accuracy is defined as the negative of the absolute value of the difference between the actual weight and the weight specified by the buyer.

> 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 [see Appendix Figure B.4]. They will have two new pieces and give the whole problem to an 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). In addition, they show the challenges of cross-border sales when, among other things, there are language barriers (both Hamis and the client quoted above communicate in English, which is not the native language of employees in either firm).

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.[49] 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 seven times, with the average firm visited 10 times. Each 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. During a visit, the intermediary discussed production techniques to achieve higher quality along these dimensions; firms reported that 91.7 percent of discussions about a particular dimension involved the intermediary providing "information on techniques to improve quality" (as opposed to only pointing out flaws). Appendix Table B.4 presents more detailed summary statistics of this dataset.

We examine if genuine knowledge was imparted on these visits as follows. We match the dataset 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 takeup firms registered larger increases relative to baseline in the particular quality dimensions that they discussed with Hamis. To perform this test, we once more stack the quality measures, indexed by d, and run the following cross-sectional regression:

$$Quality_{id} = \alpha_3 + \beta_3 Takeup_i \times \mathbf{1}[Talked\_About\_Dimension]_{id} + \gamma_3 Quality_{id0} + \delta_i + \delta_d + \varepsilon_{id}. \quad (16)$$

We include firm fixed effects δ_i so that we explicitly compare across quality dimensions d within a firm. We also include quality metric fixed effects δ_d to control for different means across dimensions.[50] This regression asks if improvements along the quality dimensions discussed were larger than improvements along the dimensions that were not discussed. A significant β_3 coefficient is supportive of the presence of knowledge transfers (and inconsistent with a simple movement along the PPF, where quality would be independent of knowledge flows).

[49] Unfortunately, due to a miscommunication, Hamis Carpets failed to record the date of these interactions so we are only able to examine cumulative interactions.

[50] Note that we do not need to include additional controls for cumulative production since cumulative production varies only at the firm level and we include firm fixed effects (and similarly we do not include the main effect of takeup).

Table 13 reports the results. Using either sample, we find a positive and statistically significant association between changes in quality and whether the intermediary discussed that quality dimension with the firm. Columns 3-6 re-run (16) controlling for rug specifications or using specification-adjusted quality metrics and the results are unchanged. These results support the hypothesis that knowledge is transferred from the intermediary to the firm.

We provide an additional piece of evidence that suggests our results are not driven entirely by learning-by-doing. Under learning-by-doing we would expect firms who 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.[51]

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 fact that firms report that 91.7 percent of discussions touched on techniques is compelling. Additionally, the intermediary has provided us with multiple examples of production technique improvements discussed with firms. 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.[52]

6.5 Step 5: Ruling Out Alternative Hypotheses

The previous four steps, in combination, provide strong evidence that exporting raises the technical efficiency of firms. In this final step, we rule out alternative explanations that could explain the patterns in the data.

There are two 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 the other steps. Treatment firms score higher quality metrics and higher productivity once we condition on rug specifications and produce at a higher quality when they make identical domestic rugs (Steps 1 and 2). Second, inconsistent with Step 4, information flows should be unrelated to quality changes if firms simply had to pay an adjustment cost to raise quality.

A second closely-related hypothesis is that the opportunity to export raised the returns to investments that raise the efficiency parameters (and hence raise quality and measured productivity). 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

[51]We find a coefficient (standard error) on the interaction of -0.15 (0.12) for Sample 2 and 0.05 (0.04) for the Joint Sample. As in previous specifications, we also include round and strata fixed effects in this regression.

[52]Relatedly, 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 upon request.

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. Additionally, we asked treatment firms about the extent to which they practiced weaving techniques, and none report ever practicing techniques.

A more complicated 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 uncorrelated with the number of hours the firm was visited by the intermediary.[53] Instead, the knowledge transfers occurring during these interactions appear to be just that: flows of information that are not priced, which is similar to the types of information flows described in the classic learning-by-exporting literature (e.g., Clerides et al. 1998).

7 A Coda on Failures in the Market for Information

The analysis in the previous section establishes that knowledge is transferred through exporting and that there are high returns to this knowledge in our setting. Given these findings, a puzzle that remains is why the knowledge did not spread quickly to the control group and eliminate any experimental treatment effect. There are two obvious possibilities. The first is that this knowledge is not sufficiently useful outside of the export market to justify the costs to acquiring it (and that the fixed costs of accessing export markets are too large to be optimal for a second intermediary or single firm to pay). The second is that there are failures in the market for information.

While identifying potential market failures is not the focus of this paper, nor the learning-by-exporting literature more generally, the question is of natural interest to both policymakers and economists. Rigorously identifying the sources of market failures would, of course, necessitate designing a different experiment. However, we can provide some evidence that suggest the latter hypothesis is the more likely one.

The first piece of evidence is that the benefits of the knowledge transfer, even in the domestic market, far exceed the costs of providing this knowledge (inconsistent with the first hypothesis). To determine the value of improving quality for a firm that sells to the domestic market, we regress profits per hour for control firms (i.e. those selling to the domestic market) on our quality metrics.[54] Combining these estimates of the domestic returns to quality with the treatment effects estimated in Section 6.2, the benefit for control firms to move to the quality levels achieved by treatment firms would be 10.4 percent higher profits on the domestic market. Obtaining this knowledge would raise average firm profits by LE1,091 per year (a lifetime net present value of LE10,910 ($1,772) using a discount rate of 10 percent). Using the wage the intermediary pays their employ-

[53]A regression among takeup firms of the log price received on log total hours of visits by the intermediary and specification controls gives a positive and insignificant coefficient of 0.08 (s.e. of 0.16) for Sample 2 and 0.05 (s.e. 0.09) for the Joint Sample.

[54]This is a regression of log profits per hour on the 9 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 zero is rejected at the 5.3 percent level. The regression is reported in Appendix Table B.5.

ees who visit the firms, we estimate that the cost of providing this training is only LE103 ($16) if we conservatively assume that all the time spent on firm visits is spent discussing techniques.

The second piece of evidence comes from explicitly looking for knowledge spillovers. If a market for knowledge transfers exists, either explicit or implicit, knowledge is most likely to flow between geographically-proximate treatment and control firms where the costs are presumably lowest. We investigate this question by regressing our main outcomes for control firms on the sum of the inverse distance to treatment firms:

$$y_{it} = \alpha_4 + \beta_4 \sum_{i'} \frac{1[i' \in Treatment]}{dist_{ii'}} + \beta_5 \sum_{i'} \frac{1[i' \in Treatment]}{(dist_{ii'})^2} + \gamma_4 y_{i0} + \delta_s + \tau_t + \varepsilon_{it}, \quad \text{for } i \in Control.$$

We run this specification for the four key outcome variables: profits, specification-adjusted productivity and (stacked) quality, and (stacked) quality scores from the lab. If geographic spillovers are present, we expect better outcomes for firms with higher values for the inverse distance sums who happen to be located closer to more treatment firms. We find little evidence of such spillovers, with none of the coefficients, or the marginal effect $\beta_4 + 2\beta_5 \sum_{i'} \frac{1[i' \in Treatment]}{dist_{ii'}}$, significantly different from zero (see Appendix Table B.6).

Taken together, it appears that either control firms do not know that they are lacking this valuable knowledge, treatment firms do not realize that the knowledge they posses is valuable, or market failures are preventing a market for training services being established in Fowa. This latter possibility echoes the work of Bloom et al. (2013) which discusses potential failures in the market for consulting services for firms in developing countries.

8 Conclusion

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

We find that operating profits for treatment firms increase 15-25 percent 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, indicating that foreign buyers demand higher quality products that take longer to manufacture. The quality upgrading we observe may or may not come about through learning-by-exporting, export induced improvements in technical efficiency biased towards the manufacture of high-quality rugs. For example, firms may have always known how to produce high-quality products but optimally chose not to because domestic buyers were unwilling to pay for them.

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 information flowing between foreign buyers and the intermediary, and between the intermediary and the producers; analyzing the latter flows shows that quality levels responded most along the particular dimensions discussed. Fifth, we find no evidence of investments, firms paying monetary adjustment costs or firms implicitly paying the intermediary for the information they receive.

Taken together, the evidence indicates that learning-by-exporting is present in our data and that the learning occurs, at least in part, through information flows. Given that this learning is induced by demand for high-quality products from high-income foreign buyers, 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.

References

AHN, J., A. K. KHANDELWAL, AND S.-J. WEI (2011): "The role of intermediaries in facilitating trade," *Journal of International Economics*, 84, 73–85.

ANGELUCCI, M., D. KARLAN, AND J. ZINMAN (2015): "Microcredit Impacts: Evidence from a Randomized Microcredit Program Placement Experiment by Compartamos Banco," *American Economic Journal: Applied Economics*, 7, 151–82.

ARTOPOULOS, A., D. FRIEL, AND J. HALLAK (2013): "Export emergence of differentiated goods from developing countries: Export pioneers and business practices in Argentina," *Journal of Development Economics*, 105, 19–35.

AW, B. Y., M. J. ROBERTS, AND Y. XU (2011): "R&D Investment, Exporting, and Productivity Dynamics," *American Economic Review*, 101, 1312–44.

BALTAGI, B. H. AND D. LI (2002): "Series Estimation of Partially Linear Panel Data Models with Fixed Effects," *Annals of Economics and Finance*, 3, 103–116.

BANERJEE, A. V. (2013): "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, 487–519.

BASTOS, P., J. SILVA, AND E. VERHOOGEN (2014): "Export Destinations and Input Prices," Unpub. paper.

BERNARD, A. B. AND B. JENSEN (1999): "Exceptional exporter performance: cause, effect, or both?" *Journal of International Economics*, 47, 1–25.

BLOOM, N., B. EIFERT, A. MAHAJAN, D. MCKENZIE, AND J. ROBERTS (2013): "Does Management Matter? Evidence from India," *The Quarterly Journal of Economics*, 128, 1–51.

BLOOM, N. AND J. VAN REENEN (2007): "Measuring and Explaining Management Practices Across Firms and Countries," *The Quarterly Journal of Economics*, 122, 1351–1408.

——— (2010): "Why Do Management Practices Differ across Firms and Countries?" *Journal of Economic Perspectives*, 24, 203–24.

BRAMBILLA, I., D. LEDERMAN, AND G. PORTO (2012): "Exports, Export Destinations, and Skills," *American Economic Review*, 102, 3406–38.

CLERIDES, S. K., S. LACH, AND J. R. TYBOUT (1998): "Is Learning By Exporting Important? Micro-Dynamic Evidence From Colombia, Mexico, And Morocco," *The Quarterly Journal of Economics*, 113, 903–947.

CROZET, M., K. HEAD, AND T. MAYER (2012): "Quality Sorting and Trade: Firm-level Evidence for French Wine," *Review of Economic Studies*, 79, 609–644.

DE LOECKER, J. (2007): "Do exports generate higher productivity? Evidence from Slovenia," *Journal of International Economics*, 73, 69–98.

——— (2011): "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 79, 1407–1451.

——— (2013): "Detecting Learning by Exporting," *American Economic Journal: Microeconomics*, 5, 1–21.

DE LOECKER, J. AND P. GOLDBERG (2014): "Firm Performance in a Global Market," *The Annual Review of Economics*.

DE LOECKER, J., P. GOLDBERG, A. K. KHANDELWAL, AND N. PAVCNIK (2014): "Prices, Markups and Trade Reform," NBER Working Paper 17925.

DE MEL, S., D. MCKENZIE, AND C. WOODRUFF (2008): "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 123, 1329–1372.

——— (2009): "Measuring microenterprise profits: Must we ask how the sausage is made?" *Journal of Development Economics*, 88, 19–31.

——— (2010): "Wage Subsidies for Microenterprises," *American Economic Review*, 100, 614–18.

——— (2014): "Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka," *Journal of Development Economics*, 106, 199–210.

EATON, J., M. ESLAVA, D. JINKINS, C. KRIZAN, AND J. TYBOUT (2013): "A Search and Learning Model of Export Dynamics," Tech. rep., Penn State.

FEENSTRA, R. C. AND G. H. HANSON (2004): "Intermediaries in Entrepot Trade: Hong Kong Re-Exports of Chinese Goods," *Journal of Economics & Management Strategy, Wiley Blackwell*, 13, 3–35.

FEENSTRA, R. C. AND J. ROMALIS (2014): "International Prices and Endogenous Quality," *The Quarterly Journal of Economics*, 129, 477–527.

FOSTER, L., J. HALTIWANGER, AND C. SYVERSON (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98, 394–425.

HALL, R. AND C. JONES (1999): "Why Do Some Countries Produce So Much More Output Per Worker Than Others?" *The Quarterly Journal of Economics*, 114, 83–116.

HALLAK, J. (2006): "Product quality and the direction of trade," *Journal of International Economics*, 68, 238–265.

HALLAK, J. AND P. K. SCHOTT (2011): "Estimating Cross-Country Differences in Product Quality," *The Quarterly Journal of Economics*, 126, 417–474.

HALLAK, J. AND J. SIVADASAN (2013): "Product and process productivity: Implications for quality choice and conditional exporter premia," *Journal of International Economics*, 91, 53 – 67.

HALLAK, J. C. (2010): "A Product-Quality View of the Linder Hypothesis," *The Review of Economics and Statistics*, 92, 453–466.

HARRISON, A. AND A. RODRIGUEZ-CLARE (2010): *Trade, Foreign Investment, and Industrial Policy for Developing Countries*, Elsevier, vol. 5 of *Handbook of Development Economics*, chap. 0, 4039–4214.

KARLAN, D. AND M. VALDIVIA (2011): "Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions," *The Review of Economics and Statistics*, 93, 510–527.

KELLER, W. (2004): "International Technology Diffusion," *Journal of Economic Literature*, 42, 752–782.

KHANDELWAL, A. (2010): "The Long and Short (of) Quality Ladders," *The Review of Economic Studies*, 77, 1450–1476.

KLING, J. R., J. B. LIEBMAN, AND L. F. KATZ (2007): "Experimental analysis of neighborhood effects," *Econometrica*, 75, 83–119.

KRUGMAN, P. R. (1979): "Increasing returns, monopolistic competition, and international trade," *Journal of International Economics*, 9, 469–479.

KUGLER, M. AND E. VERHOOGEN (2012): "Prices, Plant Size, and Product Quality," *Review of Economic Studies*, 79, 307–339.

LEVITT, S. D., J. A. LIST, AND C. SYVERSON (2013): "Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant," *Journal of Political Economy*, 121, 643 – 681.

MANOVA, K. AND Z. ZHANG (2012): "Export Prices Across Firms and Destinations," *The Quarterly Journal of Economics*, 127, 379–436.

MARIN, A. G. AND N. VOIGTLANDER (2013): "Exporting and Plant-Level Efficiency Gains: It's in the Measure," NBER Working Paper 19033.

MCKENZIE, D. AND C. WOODRUFF (2013): "What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?" *The World Bank Research Observer*.

MELITZ, M. J. (2003): "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 71, 1695–1725.

MELITZ, M. J. AND S. J. REDDING (2014): "Heterogeneous Firms and Trade," in *Handbook of International Economics*, ed. by K. R. Elhanan Helpman and G. Gopinath, Elsevier, vol. 4 of *Handbook of International Economics*, 1 – 54.

PARK, A., D. YANG, X. S. SHI, AND Y. JIANG (2010): "Exporting and Firm Performance: Chinese Exporters and the Asian Financial Crisis," *The Review of Economics and Statistics*, 92, 822–842.

RAUCH, J. E. (1999): "Networks Versus Markets in International Trade," *Journal of International Economics*, 48, 7 – 35.

SCHOTT, P. K. (2004): "Across-Product Versus Within-Product Specialization in International Trade," *The Quarterly Journal of Economics*, 119, 647–678.

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

VERHOOGEN, E. A. (2008): "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 123, 489–530.

WAGNER, J. (2007): "Exports and Productivity: A Survey of the Evidence from Firm-level Data," *World Economy*, 30, 60–82.

Appendix A Measuring Productivity

This appendix discusses the measurement of productivity in our setting. One of the key challenges with standard productivity analysis is the lack of firm-specific input and output prices which introduces biases in estimates of productivity (de Loecker and Goldberg, 2014). Even in the instances where material input and output prices may be observed, researchers almost never observe the user cost of capital and typically have noisy measures of capital (e.g., book value). We avoid these measurement issues because we observe input and output quantities, and the number of looms used by firms.[1] Rug specifications also allow us to ensure that we compare output, conditional on inputs, on equivalent goods. Moreover, since all firms produce a single product, issues that arise with multi-product firms and the divisibility of inputs are not relevant in this setting (de Loecker et al, 2014).

We consider two measures of productivity. As we describe in the text, output per hour is a meaningful measure of productivity in our setting and is directly observed in the data. For this reason, it is our benchmark measure of productivity.

A second measure relaxes the assumption that labor is the only input required for production (and has constant returns) by broadening the productivity measure to depend on labor and capital. Specifically, we run the following value-added production function:

$$\ln x_{it} = \alpha_l \ln l_{it} + \alpha_k \ln k_{it} + \mathbf{Z}_{it}' \Gamma + a_{it} + \nu_{it} \tag{A.1}$$

where x_{it} is the output (in m^2) of firm i at time t, l_{it} is total hours, k_{it} is the number of active looms, and a_{it} is firm productivity. We emphasize that there is virtually no variation in the number of looms across firms (92 percent of firms report having more than one loom), but we nevertheless

[1] Marin et al. (2013) use the methodology developed by de Loecker et al. (2014) to purge productivity measures of prices and find export-induced efficiency gains, but their learning results rely on propensity score matching techniques, which requires fully specifying the selection model, rather than on random variation that we exploit.

allow the production function to depend on capital. The vector of controls, $\mathbf{Z}_{ft}$, include rug specifications, round and strata fixed effects. The ν_{ft} is an i.i.d. error term capturing unanticipated shocks and measurement error.

Although having quantity information and rug specifications deals with measurement concerns, there is still a potential simultaneity bias in estimating (A.1) since productivity is observed by the firm, but not us. The standard approach in the literature addresses simultaneity using the control function approach developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). We assume that capital is a dynamic variable subject to adjustment costs and labor is flexible. Material demand is given by $m_{it} = f_t(a_{it}, k_{it})$ and can be inverted as $a_{it} = f_t^{-1}(m_{it}, k_{it})$. We follow the literature and assume that productivity follows a first-order Markov process. We leverage the experimental setting by estimating the production function using only the control firms. This allows us to avoid parametric (or semi-parametric) assumptions on the productivity process of treatment firms, which we argue evolves with treatment over time in potentially non-linear ways.[2] We estimate the production function 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.[3] We obtain $\alpha_l = 0.77$ (s.e. of .37) and $\alpha_k = 0.23$ (s.e. of 0.97). Given that the coefficients sum to 1, we cannot reject that there are constant returns to scale.[4]

We use these coefficients to compute (unadjusted) TFP: $a_{it} = \frac{x_{it}}{l_{it}^{0.77} k_{it}^{0.23}}$. Note that we assume that the α's are identical across treatment and control firms. We believe this is justifiable since all firms produce rugs using identical technology that has not evolved during the sample period. Moreover, since firms produce a narrowly defined product–duble rugs–our assumption is in fact weaker than most papers that estimate production functions at 2-digit or 4-digit industry classifications.

Appendix References

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

OLLEY, G. S. AND A. PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–97.

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

MARIN, A. G. AND N. VOIGTLANDER (2013): "Exporting and Plant-Level Efficiency Gains: It's in the Measure," NBER Working Paper 19033.

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

[2] See de Loecker (2013) for an extensive discussion of this point.

[3] The approach by Wooldridge (2009) addresses potential identification problems with the labor coefficient discussed by Ackerberg et al. (2006).

[4] For comparison, the OLS of (A.1) gives $\alpha_l = 0.66$ (s.e. of 0.14) and $\alpha_k = 0.15$ (s.e. of 0.05). Using the levpet command in Stata to implement the approach proposed by Levinsohn and Petrin (2003) yields $\alpha_l = 0.64$ (s.e. of 0.16) and $\alpha_k = 0.29$ (s.e. of 0.14).

Table 1: Firm Sample and Takeup Statistics

	Kasaees Orders	Duble Orders			
	Sample 1		Sample 1		Sample 2
	Kasaees Firms	Goublain Firms	Tups Firms	Duble Firms	Duble Firms
Statistic	(1)	(2)	(3)	(4)	(5)
Firms	38	103	83	79	140
Treatment firms	19	49	42	39	35
Takeup firms	5	5	8	14	32
Initial packet size (m^2)	250	110	110	110	110
Successful takeup firms	5	4	6	14	32
Average output conditional on takup (m^2)	303	586	589	778	434

Notes: Table reports statistics by firm type and sample. The 1st row displays the number of firms within each rug type and sample. The 2nd row displays the number of firms in the treatment group. The 3rd row indicates the number of firms who accepted the treatment and agreed to make rugs for export. The 4th row is the initial order size (in square meters) offered to each takeup firm. The 5th row shows the number of firms that completed the initial order successfully and received subsequent orders from Hamis. The last row indicates average output conditional on takeup.

Table 2: Survey Timeline

Survey Timeline	Sample 1	Sample 2
Baseline	July-Aug 2011	Feb-Mar 2013
Round 1	§Nov-Dec 2011	May-June 2013
Round 2	April-May 2012	Nov-Dec 2013
Round 3	Sept-Dec 2012	May-June 2014
Round 4	Mar-Apr 2013	
Round 5	July-Oct 2013	
Round 6	Jan-Mar 2014	
Quality Lab	June 2014	June 2014

Notes: Table reports the timeline for the data survey collection by sample. §Data from Round 1 for Sample 1 is unreliable and is discarded in the analysis.

Table 3: Baseline Balance

	Sample 2			Joint Sample		
	Constant	Treatment	N	Constant	Treatment	N
Panel A: Household Characteristics						
Age	50.0	2.8	139	50.7	0.9	218
	(1.1)	(2.2)		(0.9)	(1.6)	
Experience	36.3	1.9	136	37.6	0.2	213
	(1.2)	(2.5)		(1.0)	(1.7)	
Illiterate?	0.57	0.07	135	0.59	0.10	214
	(0.05)	(0.10)		(0.04)	(0.07)	
Household size	4.0	0.1	140	4.2	0.0	219
	(0.2)	(0.3)		(0.1)	(0.2)	
Panel B: Firm Characteristics						
Constructed profits from rug business	874	-35.6	139	820	131.0	218
	(46.4)	(87.4)		(42.9)	(99.6)	
Hours worked last month	268	1.3	139	247	-1.7	218
	(5.9)	(10.8)		(7.0)	(11.7)	
Number of employees	1.00	-	139	1.09	0.0	218
	-	-		(0.0)	(0.1)	
Total produced last month (m^2)	43.5	0.33	139	48.9	3.3	218
	(2.7)	(5.81)		(4.8)	(10.0)	
Ever exported?	0.16	0.03	140	0.11	0.02	219
	(0.04)	(0.08)		(0.03)	(0.05)	
Average Quality	2.57	-0.09	140	2.68	-0.13 ***	218
	(0.03)	(0.06)		(0.03)	(0.05)	
Joint F-test		0.86			0.85	
Attrition	0.04	-0.02	420	0.11	0.00	815
	(0.01)	(0.02)		(0.01)	(0.02)	

Notes: Table presents baseline balance for Sample 2 (left panel) and the Joint Sample (right panel). Each row is a regression of the variable on a constant, treatment dummy and strata fixed effects. The 2nd to last row reports the F-test for a test of joint signficance of the baseline variables. Real constructed profits for the Joint Sample are winsorized at the 2.5th and 97.5th percentile to trim outliers (without winsorizing, the sample still remains statistically balanced between treatment and control groups). The final row reports average attrition across all survey rounds. Significance * .10; ** .05; *** .01.

Table 4: Impact of Intervention on Firms Knowingly Exporting

	Sample 2		Joint Sample	
	Indicator if ever exported		Indicator if ever exported	
	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)
Treatment	0.68 ***		0.55 ***	
	(0.07)		(0.06)	
Takeup		0.75 ***		0.76 ***
		(0.07)		(0.07)
R-squared	0.45	0.49	0.33	0.45
Observations	132	132	191	191

Notes: Table regresses an indicator for if a firm has ever knowingly produced rugs for export markets on indicators of 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.

Table 5: Impact of Exporting on Firm Profits

Panel A: Profits for Sample 2

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.25 ***		0.23 ***		0.24 ***		0.36 ***	
	(.06)		(.05)		(.05)		(.10)	
Takeup		0.30 ***		0.28 ***		0.29 ***		0.44 ***
		(.07)		(.06)		(.07)		(.12)
R-squared	0.29	0.30	0.27	0.28	0.28	0.28	0.33	0.34
Observations	375	375	375	375	375	375	373	373

Panel B: Profits for Joint Sample

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.26 ***		0.21 ***		0.19 ***		0.37 ***	
	(.05)		(.06)		(.06)		(.11)	
Takeup		0.42 ***		0.37 ***		0.32 ***		0.68 ***
		(.08)		(.10)		(.09)		(.19)
R-squared	0.21	0.22	0.16	0.18	0.19	0.22	0.19	0.19
Observations	573	573	644	644	655	655	687	687

Panel C: Profit per Owner Hour for Sample 2

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.17 ***		0.15 ***		0.16 ***		0.21 ***	
	(.05)		(.05)		(.05)		(.06)	
Takeup		0.20 ***		0.18 ***		0.19 ***		0.26 ***
		(.06)		(.06)		(.06)		(.08)
R-squared	0.20	0.21	0.20	0.20	0.19	0.19	0.32	0.31
Observations	375	375	375	375	375	375	373	373

Panel D: Profit per Owner Hour for Joint Sample

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.19 ***		0.16 ***		0.15 ***		0.25 ***	
	(.05)		(.05)		(.05)		(.06)	
Takeup		0.31 ***		0.28 ***		0.25 ***		0.46 ***
		(.08)		(.09)		(.09)		(.12)
R-squared	0.17	0.17	0.15	0.15	0.16	0.17	0.23	0.21
Observations	573	573	644	644	654	654	687	687

Notes: Table reports treatment effects on different real profit measures, all measured in logs. See text for details regarding each measure. Panel A runs ITT and TOT regressions on Sample 2. Panel B reports the analysis using the Joint Sample. Panels C and D report the analogous regressions using profits per hour as the dependent variable. 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.

Table 6: Sources of Changes to Firm Profits

Panel A: Sample 2

	Output Prices (LE/m^2)		Output (m^2)		Hours Worked		Warp Thread Ball (kg)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.46 ***		-0.22 **		0.08 ***		0.13 **	
	(.10)		(.09)		(.02)		(.05)	
Takeup		0.56 ***		-0.27 ***		0.10 ***		0.15 **
		(.12)		(.10)		(.03)		(.06)
R-squared	0.27	0.28	0.19	0.19	0.12	0.12	0.20	0.20
Observations	376	376	375	375	375	375	377	377

Panel B: Joint Sample

	Output Prices (LE/m^2)		Output (m^2)		Hours Worked		Warp Thread Ball (kg)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.49 ***		-0.26 ***		0.05 **		0.15 ***	
	(.09)		(.09)		(.02)		(.05)	
Takeup		0.85 ***		-0.47 ***		0.08 **		0.25 ***
		(.16)		(.17)		(.04)		(.08)
R-squared	0.25	0.23	0.24	0.22	0.12	0.13	0.24	0.24
Observations	665	665	676	676	678	678	600	600

Notes: Table reports treatment effects on real prices, output, hours worked and size of the warp thread ball, all measured in logs. The TOT regression instruments takeup with treatment. 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.

Table 7: Impacts on Input Prices and Quantities

Panel A: Sample 2

	Weft Thread Price		Warp Thread Price		Weft Thread Quantity (g)		Warp Thread Quantity (g)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.23 ***		-0.03		-0.13		0.08	
	(.04)		(.03)		(.09)		(.09)	
Takeup		0.29 ***		-0.04		-0.16		0.10
		(.05)		(.04)		(.10)		(.11)
R-squared	0.12	0.14	0.14	0.14	0.17	0.17	0.14	0.14
Observations	376	376	376	376	375	375	375	375

Panel B: Joint Sample

	Weft Thread Price		Warp Thread Price		Weft Thread Quantity (g)		Warp Thread Quantity (g)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.20 ***		-0.04		-0.19 **		0.08	
	(.06)		(.03)		(.10)		(.09)	
Takeup		0.33 ***		-0.07		-0.34 **		0.10
		(.10)		(.06)		(.17)		(.11)
R-squared	0.22	0.24	0.27	0.27	0.23	0.22	0.14	0.14
Observations	564	564	685	685	677	677	375	375

Notes: Table reports treatment effects on real input price and input quantities, all measured in logs. The TOT regression instruments takeup with treatment. 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.

Table 8: Impact of Exporting on Quality Levels

Panel A: Quality Metrics

	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
Corners	1.38 ***	1.69 ***	1.11 ***	1.70 ***
	(0.13)	(0.08)	(0.12)	(0.11)
Waviness	1.36 ***	1.66 ***	1.10 ***	1.68 ***
	(0.13)	(0.08)	(0.12)	(0.10)
Weight	1.32 ***	1.60 ***	1.07 ***	1.63 ***
	(0.12)	(0.09)	(0.11)	(0.11)
Touch	0.54 ***	0.65 ***	0.40 ***	0.66 ***
	(0.08)	(0.06)	(0.06)	(0.07)
Packedness	1.38 ***	1.68 ***	0.89 ***	1.59 ***
	(0.12)	(0.08)	(0.11)	(0.12)
Warp Thread Tightness	1.24 ***	1.51 ***	0.83 ***	1.49 ***
	(0.12)	(0.09)	(0.10)	(0.12)
Firmness	1.43 ***	1.75 ***	0.87 ***	1.60 ***
	(0.13)	(0.08)	(0.11)	(0.12)
Design Accuracy	1.22 ***	1.48 ***	0.79 ***	1.41 ***
	(0.12)	(0.10)	(0.10)	(0.12)
Warp Thread Packedness	1.33 ***	1.64 ***	1.07 ***	1.65 ***
	(0.13)	(0.09)	(0.11)	(0.11)
Inputs	1.37 ***	1.66 ***	0.89 ***	1.62 ***
	(0.11)	(0.09)	(0.10)	(0.12)
Loom	0.04	0.05	0.03	0.05
	(0.04)	(0.04)	(0.02)	(0.04)
R-squared	0.57	0.66	0.44	0.60
Observations	4,120	4,120	6,885	6,885

Panel B: Stacked Quality Metrics

	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
Stacked Quality Metrics	1.14 ***	1.39 ***	0.79 ***	1.35 ***
	(0.10)	(0.06)	(0.09)	(0.08)
R-squared	0.52	0.60	0.39	0.54
Observations	4,120	4,120	6,885	6,885

Notes: Panel A stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality metric indicator, so each coefficient is the differential impact for each metric between treatment and control. The TOT instruments takeup (interacted with quality metric) with treatment (also interacted with quality metric). Each regression includes baseline values of the quality metric, strata and round fixed effects, and each of these controls is interacted with quality metric indicators. Standard errors are clustered by firm. Panel B constrains the ITT and TOT to be the same across quality metrics; these regressions include baseline values, strata and round fixed effects with standard errors clustered by firm. Significance * .10; ** .05; *** .01.

Table 9: Impact of Exporting on Productivity

	Panel A: Output Per Hour			
	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
Output Per Hour	-0.24 ***	-0.29 ***	-0.24 ***	-0.42 ***
	(0.09)	(0.10)	(0.09)	(0.16)
R-squared	0.18	0.20	0.18	0.16
Observations	376	376	687	687

	Panel B: TFP			
	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
TFP	-0.29 ***	-0.35 ***	-0.29 ***	-0.51 ***
	(0.09)	(0.11)	(0.09)	(0.16)
R-squared	0.19	0.20	0.26	0.24
Observations	375	375	674	674

Notes: Table reports treatment effects on the two productivity measures: (log) output per hour and (log) TFP. See Appendix A for the methodology used to obtain TFP. The TOT specifications instrument takeup with treatment. Regressions control for baseline values of the variable, round and strata fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 10: Conditional Quality and Productivity

Panel A: Stacked Quality Metrics

	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
Treatment	0.53 ***		0.31 ***	
	(0.10)		(0.04)	
Takeup		0.83 ***		0.78 ***
		(0.09)		(0.08)
(log) Thread quantity	0.06	0.09	0.03	0.02
	(0.10)	(0.07)	(0.06)	(0.05)
Difficulty Control	0.43 ***	0.34 ***	0.47 ***	0.34 ***
	(0.04)	(0.04)	(0.02)	(0.03)
(log) # colors	0.01	-0.01	0.03 *	0.00
	(0.02)	(0.02)	(0.01)	(0.01)
Low-market Segment	-0.16 ***	-0.09 **	-0.20 ***	-0.08 ***
	(0.04)	(0.04)	(0.03)	(0.03)
Mid-Market Segment	-0.11 **	-0.03	-0.19 ***	-0.06
	(0.05)	(0.05)	(0.04)	(0.04)
Rug Type FEs	yes	yes	yes	yes
R-squared	0.66	0.69	0.64	0.67
Observations	4,076	4,076	6,820	6,820

	Panel B: Productivity: Output per Hour				Panel C: Productivity: TFP			
	Sample 2		Joint Sample		Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.28 ***		0.15 **		0.24 **		0.12 *	
	(0.10)		(0.08)		(0.10)		(0.07)	
Takeup		0.43 ***		0.39 **		0.38 **		0.29 *
		(0.16)		(0.18)		(0.15)		(0.17)
(log) Thread quantity	-0.45 **	-0.43 **	-0.10	-0.11	-0.36 **	-0.34 **	-0.03	-0.04
	(0.19)	(0.19)	(0.13)	(0.12)	(0.18)	(0.17)	(0.12)	(0.12)
Difficulty Control	-0.12 **	-0.16 ***	-0.16 ***	-0.22 ***	-0.14 ***	-0.18 ***	-0.18 ***	-0.23 ***
	(0.05)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)	(0.05)
(log) # colors	-0.05	-0.06	-0.04	-0.06 **	-0.04	-0.06	-0.05 **	-0.06 **
	(0.05)	(0.05)	(0.03)	(0.03)	(0.04)	(0.04)	(0.02)	(0.02)
Low-market Segment	0.53 ***	0.57 ***	0.41 ***	0.47 ***	0.54 ***	0.57 ***	0.43 ***	0.47 ***
	(0.10)	(0.11)	(0.09)	(0.10)	(0.10)	(0.10)	(0.08)	(0.09)
Mid-Market Segment	0.30 ***	0.34 ***	0.26 ***	0.33 ***	0.32 ***	0.36 ***	0.25 ***	0.30 ***
	(0.10)	(0.11)	(0.08)	(0.09)	(0.10)	(0.10)	(0.08)	(0.09)
Rug Type FEs	yes	yes	yes	yes	yes	yes	yes	yes
R-squared	0.53	0.53	0.55	0.56	0.54	0.54	0.61	0.62
Observations	371	371	673	673	370	370	660	660

Notes: Table reports treatment effects on the stacked quality measures, and the two productivity measures: (log) output per hour and (log) TFP. See Appendix A for the methodology used to obtain TFP. The TOT specifications instrument takeup with treatment. There are 7 rug types fixed effects. In addition to the controls displayed in the table, the regressions also control for baseline values of the variable, round and strata and rug type fixed effects. The regressions in Panel A control for quality metric fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 11: Specification-Adjusted Quality and Productivity

	Panel A: Stacked Adjusted Quality Metrics			
	Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)
Treatment	0.61 ***		0.42 ***	
	(0.06)		(0.05)	
Takeup		0.75 ***		0.72 ***
		(0.04)		(0.04)
R-squared	0.26	0.32	0.18	0.27
Observations	4,076	4,076	6,860	6,860

	Panel B: Adjusted Output per Hour				Panel C: Adjusted TFP			
	Sample 2		Joint Sample		Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment	0.32 ***		0.16 **		0.32 ***		0.18 **	
	(0.07)		(0.08)		(0.07)		(0.08)	
Takeup		0.39 ***		0.30 **		0.39 ***		0.32 **
		(0.08)		(0.13)		(0.08)		(0.13)
R-squared	0.15	0.14	0.06	0.09	0.15	0.16	0.07	0.11
Observations	371	371	678	678	370	370	671	671

Notes: Table reports treatment effects on the stacked adjusted quality metrics, and the two adjusted productivity measures: (log) output per hour and (log) TFP. The adjustment uses the two-stage procedure described in Section 6.1, and Appendix A describes the methodology used to obtain TFP. The TOT specifications instrument takeup with treatment. Regressions control for baseline values of the variable, round and strata fixed effects. The regressions in Panel A control for quality metric fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 12: Quality and Productivity on Identical Domestic Rugs

Panel A: Quality Metrics

	Sample 2				Joint Sample			
	Master Artisan		Professor		Master Artisan		Professor	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Corners	0.89 ***	0.96 ***	0.40 **	0.42 **	0.72 ***	1.05 ***	0.29 **	0.45 **
	(0.16)	(0.15)	(0.17)	(0.18)	(0.14)	(0.17)	(0.13)	(0.18)
Waviness	0.72 ***	0.78 ***	0.29 *	0.30 *	0.55 ***	0.83 ***	0.25 **	0.36 **
	(0.17)	(0.16)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
Weight	0.85 ***	0.96 ***	0.62 **	0.74 ***	0.62 ***	0.91 ***	0.58 ***	1.01 ***
	(0.16)	(0.16)	(0.24)	(0.26)	(0.13)	(0.17)	(0.17)	(0.27)
Packedness	1.09 ***	1.19 ***	0.43 ***	0.48 ***	0.77 ***	1.10 ***	0.28 **	0.43 ***
	(0.15)	(0.15)	(0.14)	(0.15)	(0.13)	(0.16)	(0.11)	(0.16)
Touch	0.73 ***	0.80 ***	0.43 ***	0.47 ***	0.52 ***	0.79 ***	0.36 ***	0.53 ***
	(0.13)	(0.13)	(0.15)	(0.16)	(0.11)	(0.14)	(0.12)	(0.17)
Warp Thread Tightness	0.66 ***	0.71 ***	0.43 ***	0.49 ***	0.51 ***	0.74 ***	0.25 **	0.39 **
	(0.10)	(0.10)	(0.15)	(0.15)	(0.09)	(0.12)	(0.12)	(0.17)
Firmness	1.04 ***	1.13 ***	0.438 ***	0.49 ***	0.71 ***	1.01 ***	0.29 **	0.43 **
	(0.14)	(0.14)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
Design Accuracy	0.68 ***	0.79 ***	0.45 ***	0.48 ***	0.53 ***	0.83 ***	0.27 **	0.39 **
	(0.14)	(0.15)	(0.12)	(0.14)	(0.11)	(0.16)	(0.11)	(0.16)
Warp Thread Packedness	1.12 ***	1.20 ***	0.57 ***	0.65 ***	0.87 ***	1.28 ***	0.39 ***	0.62 ***
	(0.16)	(0.16)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
R-squared	0.31	0.34	0.10	0.08	0.21	0.32	0.11	0.11
Observations	1,087	1,087	1,078	1,078	1,680	1,680	1,667	1,667

Panel B: Stacked Quality Metrics

	Sample 2				Joint Sample			
	Master Artisan		Professor		Master Artisan		Professor	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Stacked Quality Metric	0.87 ***	0.95 ***	0.45 ***	0.50 ***	0.64 ***	0.94 ***	0.33 ***	0.48 ***
	(0.12)	(0.11)	(0.12)	(0.13)	(0.10)	(0.12)	(0.10)	(0.13)
R-squared	0.29	0.34	0.09	0.09	0.19	0.32	0.09	0.13
Observations	1,087	1,087	1,078	1,078	1,680	1,680	1,667	1,667

Panel C: Additional Quality Metrics

	ITT	TOT	ITT	TOT
	(1A)	(1B)	(3A)	(3B)
Length Accuracy	1.93 ***	1.95 ***	1.43 ***	1.97 ***
	(0.63)	(0.72)	(0.51)	(0.75)
Width Accuracy	0.43	0.47	0.17	0.26
	(0.34)	(0.38)	(0.29)	(0.45)
Weight Accuracy	93.3 ***	108.0 ***	89.1 ***	148.0 ***
	(29.1)	(30.6)	(20.3)	(32.2)
Time (in minutes)	5.52	5.40	-5.67	-8.9
	(7.4)	(7.9)	(6.6)	(9.7)
R-squared	0.83	0.83	0.84	0.82
Observations	484	484	748	748

Notes: Table reports ITT and TOT specifications using the 9 quality metrics from the quality lab. For Panel A, the ITT reports the interaction of the quality metric with a treatment dummy, and the TOT reports the interaction of the quality metric with takeup, where takeup is instrumented with the quality metric interacted with treatment. Panel B reports the results when the metrics are stacked. Columns 1 and 3 report scores from the master artisan. Columns 2 and 4 report scores from the Professor of Handicraft Science. Panel C reports 3 additional quality metrics and the time spent to produce the rug. All regressions include interactions of strata fixed effects with quality metric, and standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 13: Information Flows and Quality Levels

	Stacked Quality Metrics					
	No Controls		Specification Controls		Specification Adjusted	
	Sample 2	Joint Sample	Sample 2	Joint Sample	Sample 2	Joint Sample
	(1)	(2)	(3)	(4)	(5)	(6)
Takeup_i x $\{\text{Talked About Dimension}\}_{id}$	0.20 **	0.19 **	0.16 *	0.18 **	0.16 *	0.16 **
	(0.10)	(0.08)	(0.09)	(0.07)	(0.09)	(0.07)
Quality Metric FEs	yes	yes	yes	yes	yes	yes
Product characteristic controls	no	no	yes	yes	no	no
Specification-adjusted Quality Metrics	no	no	no	no	yes	yes
R-squared	0.76	0.76	0.76	0.76	0.46	0.44
Observations	1,098	1,700	1,068	1,660	1,068	1,667

Notes: Table regresses stacked quality metrics on on takeup indicator and its interaction with a dummy that takes the value 1 if the intermediary talked to the firm about the particular quality metric. Columns 3-4 control for rug specifications, and columns 5-6 use the specification-adjusted quality metrics described in the text. Regressions are run on a cross-section of firms and include baseline values, firm fixed effects and quality metric fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Figure 1: Production Technology

Figure 2: Example of Rug Specifications Provided by Potential Foreign Client

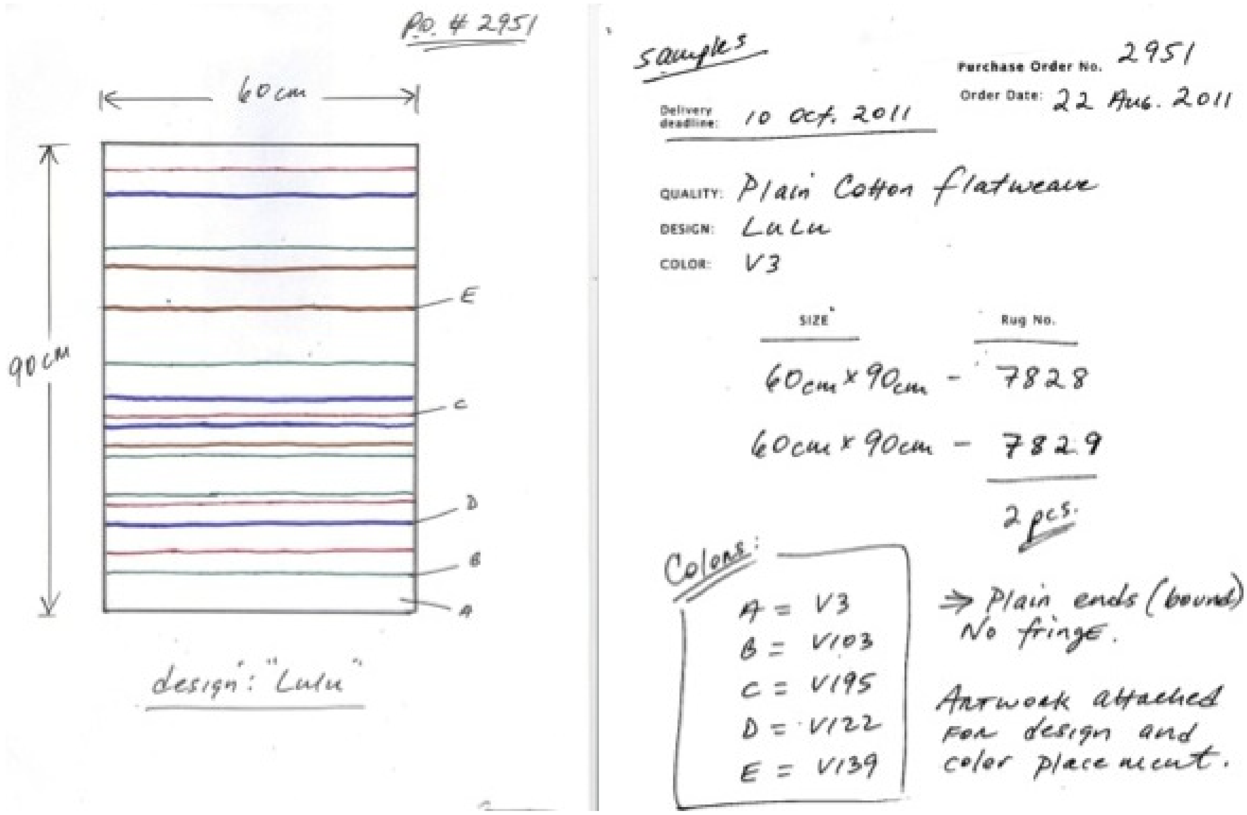


Figure 3: Example of Rugs Ordered by high-income OECD Clients

Figure 4: Cumulative Export Orders

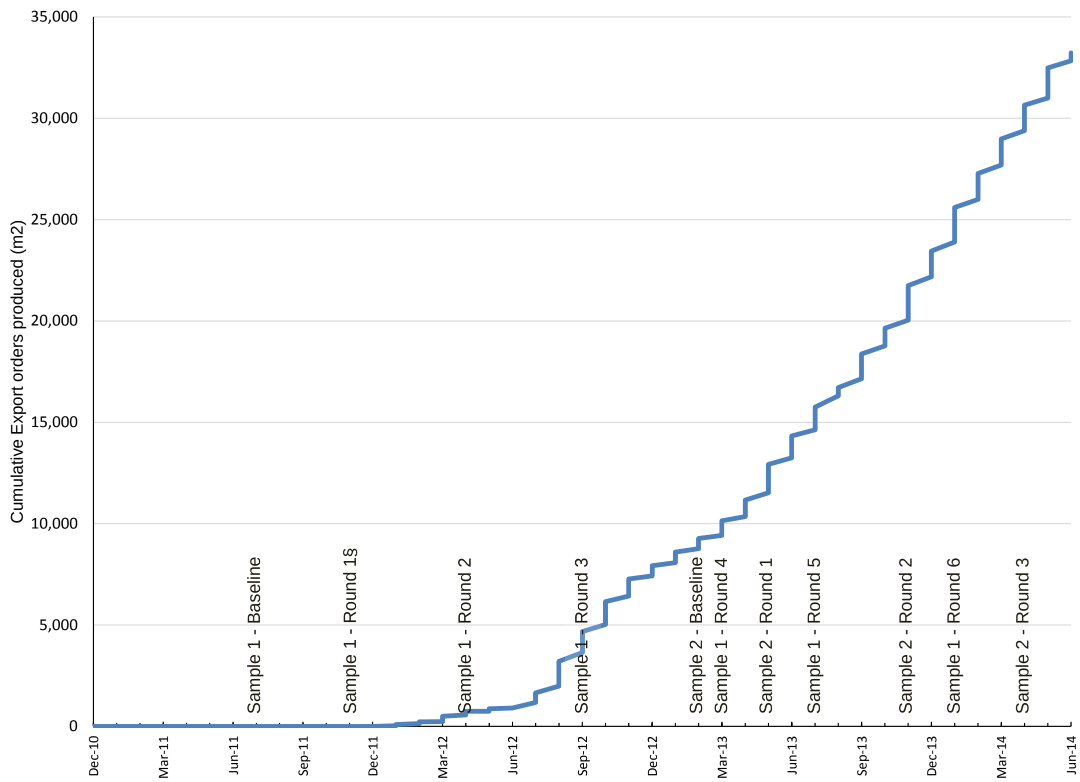


Figure 5: Overlap in Rug Specifications on Domestic and Export Orders

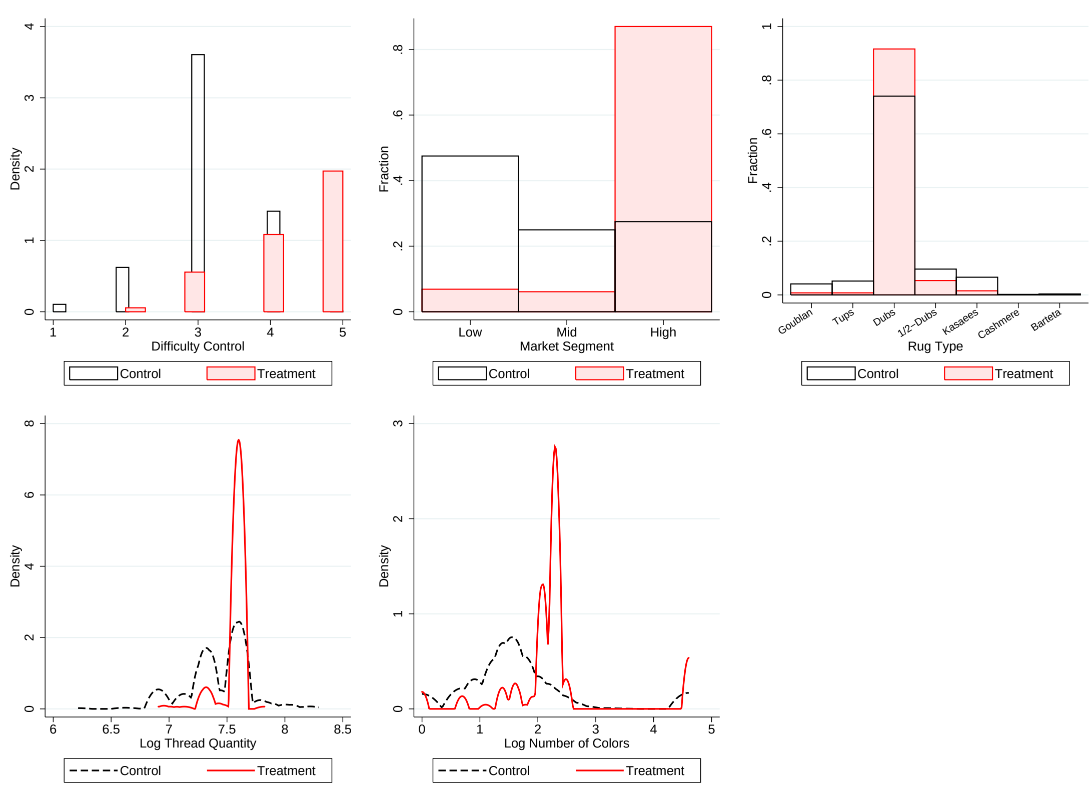


Figure 6: Learning Curves, Sample 2 Takeup Firms

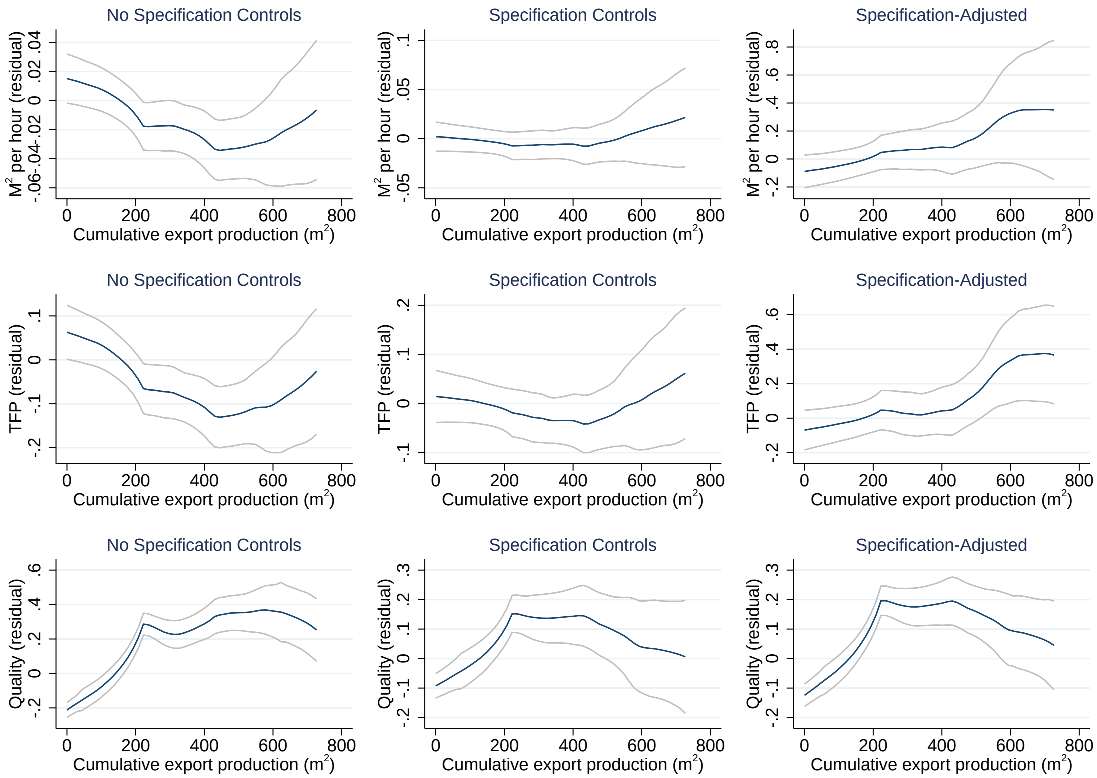
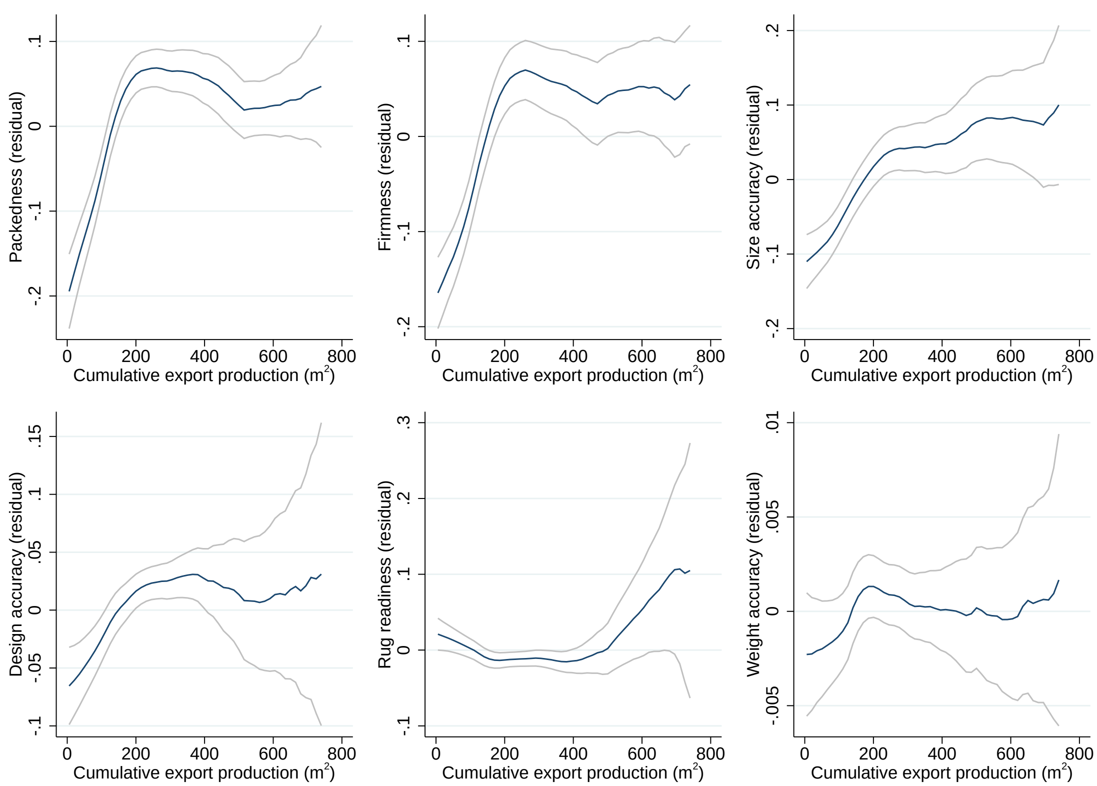


Figure 7: Learning Curves using High-Frequency Order-Book Data, Sample 2 Takeup Firms



Appendix B Tables and Figures not for Publication

Table B.1: Reasons for Refusing Treatment, Sample 1

	Goublain Firms		Tups Firms		Kasaees Firms		Duble Firms		All Firms	
Reasons for Refusal	N	%	N	%	N	%	N	%	N	%
(Agreed)	3	6	6	14	5	26	15	38	28	19
Risk relationship with current intermediary	2	4	1	2	2	11	7	18	12	8
Price was too low	2	4	1	2	2	11	3	8	9	6
Left industry or passed away	2	4	3	7	3	16	5	13	13	9
Export order not suitable rug type	39	80	30	71	6	32	7	18	82	55
Refused contact with survey team	1	2	1	2	1	5	2	5	5	3
Total	49	100	42	100	19	100	39	100	149	100

Notes: Table reports the reasons for refusing treatment orders among Sample 1 firms from the second survey round (April-May 2012). As of the second survey round, 28 firms had agreed to take orders. Since that time, an additional duble firm, two additional goublain firms and two additional tups firms have also taken orders resulting in a total of 33 Sample 1 firms takeup firms.

Table B.2: Hamis Carpets' Cost Structure

	Revenue and Expenses, per m^2	
	Domestic Orders	Export Orders
Material Expenses	30	40
Payments to Producers	25	40
Shipping Costs	0	40
Price Received	60	160
Markup	9%	33%

Notes: Table reports Hamis Carpets' cost structure on foreign and domstic rugs. Numbers reported in Egyptian Pounds per square meter.

Table B.3: Conditional and Specification-Adjusted Quality, by Metric

	Controlling for Rug Specifications				Adjusting for Rug Specifications			
	Sample 2		Joint Sample		Sample 2		Joint Sample	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Corners	0.69 ***	1.07 ***	0.47 ***	1.07 ***	0.68 ***	0.83 ***	0.53 ***	0.82 ***
	(0.14)	(0.14)	(0.09)	(0.17)	(0.08)	(0.07)	(0.07)	(0.08)
Waviness	0.61 ***	0.95 ***	0.41 ***	0.93 ***	0.57 ***	0.70 ***	0.46 ***	0.70 ***
	(0.14)	(0.15)	(0.08)	(0.14)	(0.08)	(0.07)	(0.06)	(0.07)
Weight	0.59 ***	0.92 ***	0.38 ***	0.88 ***	0.69 ***	0.85 ***	0.55 ***	0.84 ***
	(0.13)	(0.14)	(0.08)	(0.15)	(0.08)	(0.07)	(0.07)	(0.08)
Packedness	0.60 ***	0.94 ***	0.33 ***	0.86 ***	0.84 ***	1.03 ***	0.57 ***	1.03 ***
	(0.12)	(0.13)	(0.06)	(0.14)	(0.08)	(0.06)	(0.07)	(0.10)
Touch	0.27 ***	0.42 ***	0.19 ***	0.46 ***	0.49 ***	0.60 ***	0.36 ***	0.60 ***
	(0.09)	(0.11)	(0.05)	(0.09)	(0.07)	(0.06)	(0.05)	(0.07)
Warp Thread Tightness	0.46 ***	0.71 ***	0.22 ***	0.57 ***	0.57 ***	0.70 ***	0.42 ***	0.77 ***
	(0.09)	(0.10)	(0.05)	(0.10)	(0.07)	(0.07)	(0.06)	(0.10)
Firmness	0.77 ***	1.20 ***	0.39 ***	1.04 ***	1.16 ***	1.44 ***	0.67 ***	1.25 ***
	(0.14)	(0.13)	(0.06)	(0.13)	(0.11)	(0.07)	(0.09)	(0.11)
Design Accuracy	0.57 ***	0.87 ***	0.29 ***	0.76 ***	0.68 ***	0.82 ***	0.43 ***	0.79 ***
	(0.12)	(0.14)	(0.06)	(0.14)	(0.08)	(0.08)	(0.07)	(0.11)
Warp Thread Packedness	0.62 ***	0.99 ***	0.44 ***	1.02 ***	0.82 ***	1.02 ***	0.66 ***	1.04 ***
	(0.15)	(0.15)	(0.08)	(0.15)	(0.09)	(0.08)	(0.08)	(0.10)
Inputs	0.64 ***	1.00 ***	0.35 ***	0.91 ***	0.90 ***	1.10 ***	0.61 ***	1.13 ***
	(0.14)	(0.19)	(0.06)	(0.15)	(0.08)	(0.08)	(0.08)	(0.12)
Loom	0.04	0.07	0.02	0.04	0.00	0.00	0.01	0.01
	(0.04)	(0.06)	(0.02)	(0.05)	(0.02)	(0.03)	(0.02)	(0.03)
R-squared	0.74	0.77	0.72	0.76	0.36	0.43	0.26	0.37
Observations	4,076	4,076	6,820	6,820	4,076	4,076	6,830	6,830

Notes: The left panel stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality metric indicator, so each coefficient is the differential impact for each metric between treatment and control. The TOT instruments takeup (interacted with quality metric) with treatment (also interacted with quality metric). Each regression includes baseline values of the quality metric, strata and round fixed effects, and rug specification and each of these controls are interacted with quality metric indicators. The right panel uses adjusted quality metrics using the two-stage process described in Section 6.1 as the dependent variable. Significance * .10; ** .05; *** .01.

Table B.4: Summary of Information Flows

	Sample 2 (1)		Joint Sample (2)	
Number of Visits	10.1		11.0	
	(1.76)		(2.57)	
Length of Visit (in minutes)	27.8		27.6	
	(4.49)		(4.88)	
Discussed technique?	91.7%		91.6%	
	Discussed Metric?	Discusssed Technique?	Discussed Metric?	Discusssed Technique?
	(1A)	(1B)	(2A)	(2B)
Packedness	23.3%	100.0%	20.5%	100.0%
Corners	33.3%	100.0%	31.8%	100.0%
Waviness	23.3%	100.0%	20.5%	100.0%
Weight	50.0%	91.0%	54.5%	94.0%
Touch	10.0%	100.0%	11.4%	100.0%
Warp Thread Tightness	56.7%	75.0%	47.7%	79.0%
Firmness	30.0%	100.0%	31.8%	100.0%
Design Accuracy	53.3%	100.0%	50.0%	100.0%
Warp Thread Packedness	26.7%	67.0%	22.7%	75.0%
Observations	30		44	

Notes: Table summarize the visits between the intermediary with firms. All firms were visited at least 7 times, and the top panel reports the average length of visit in minutes (with standard deviations in parantheses), and the proportion of interactions overall that discuss technique, rather than simply pointing out flaws. The bottom panel reports the proportion of firms that report discussing the quality metric with the intermediary, and the proportion of firms that report discussing technique on that metric. Note that the data were collected before the final two take-up firms in Sample 2 began producing for export.

Table B.5: Quality Hedonic Regression

Quality Metric	Joint Sample (1)
Corners	0.092
	(0.067)
Waviness	-0.011
	(0.068)
Weight	-0.053
	(0.052)
Touch	0.188 **
	(0.079)
Packedness	-0.154 *
	(0.080)
Warp Thread Tightness	0.178 **
	(0.084)
Firmness	-0.134
	(0.093)
Design Accuracy	0.102 *
	(0.061)
Warp Thread Packedness	0.057
	(0.074)
P-Value of Joint F-Test	0.053
R-squared	0.584
Observations	465

Notes: Table reports the regression of profits per hour on nine quality metrics (loom and input metrics are excluded since they were not measured in Step 2 as noted in the text). The regression is run on the joint sample of control firms. The regressions include for specifications, strata and round fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table B.6: Spillovers to Control Firms

	Sample 2			
	Direct Profits	Output Per Hour	Stacked Specification-Adjusted Quality	Stacked Quality from Lab
	(1)	(2)	(3)	(4)
Sum of Inverse Distance to Treatment Firms	0.85	-0.38	0.44	0.67
	(1.21)	(0.91)	(0.48)	(2.44)
Sum of Inverse Squared Distance to Treatment Firms	-51.80	-1.64	-9.85	-49.80
	(39.27)	(37.15)	(17.41)	(77.44)
Marginal Effect	-20.92	-1.07	-3.70	-20.08
	(15.47)	(14.93)	(6.91)	(30.26)
R-Squared	0.31	0.10	0.03	0.03
Observations	272	273	3,002	773

	Joint Sample			
	Direct Profits	Output Per Hour	Stacked Specification-Adjusted Quality	Stacked Quality from Lab
	(1)	(2)	(3)	(4)
Sum of Inverse Distance to Treatment Firms	-0.89	-1.44	0.60	-0.11
	(1.35)	(1.00)	(0.41)	(2.00)
Sum of Inverse Squared Distance to Treatment Firms	0.28	48.28	-7.09	-16.90
	(43.17)	(40.32)	(15.91)	(64.35)
Marginal Effect	-0.77	19.11	-2.41	-7.26
	(17.22)	(16.36)	(6.43)	(25.55)
R-Squared	0.26	0.08	0.06	0.06
Observations	368	427	4,408	1,094

Notes: Table reports results from regressing the logged outcome variables in each column on inverse distances between control firms and all treatment firms and an inverse distance squared term (measured in meters). The third row shows the marginal effect of distance on the outcome based on the results from the regression. The stacked quality metrics in columns 3 and 4 are not logged, and those regressions include metric fixed effects. The quality lab column uses the Master Artisan's grades. Regressions include round and strata fixed effects. Columns 1-3 control for baseline values. Standard errors clustered at the firm level.

Figure B.1: Examples of Duble, Tups, Kasaees, and Goublan Rugs

Figure B.2: Learning Curves, Sample 2 Takeup Firms

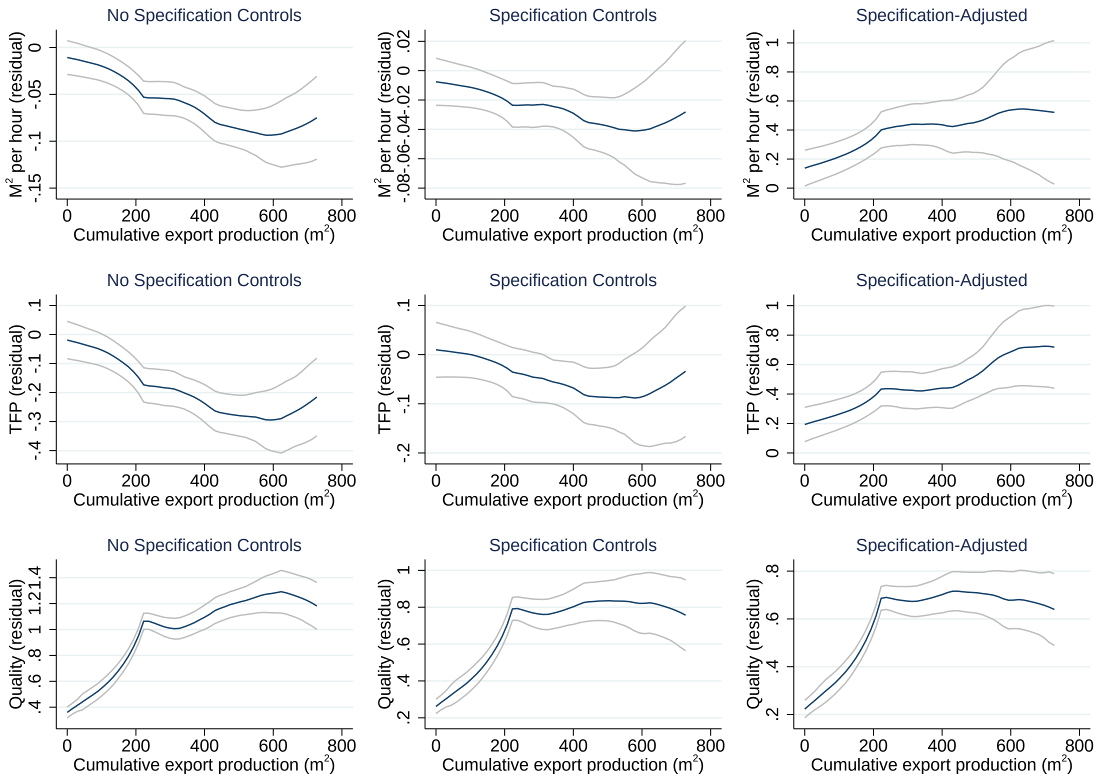


Figure B.3: Learning Curves using High-Frequency Order-Book Data, Sample 2 Takeup Firms

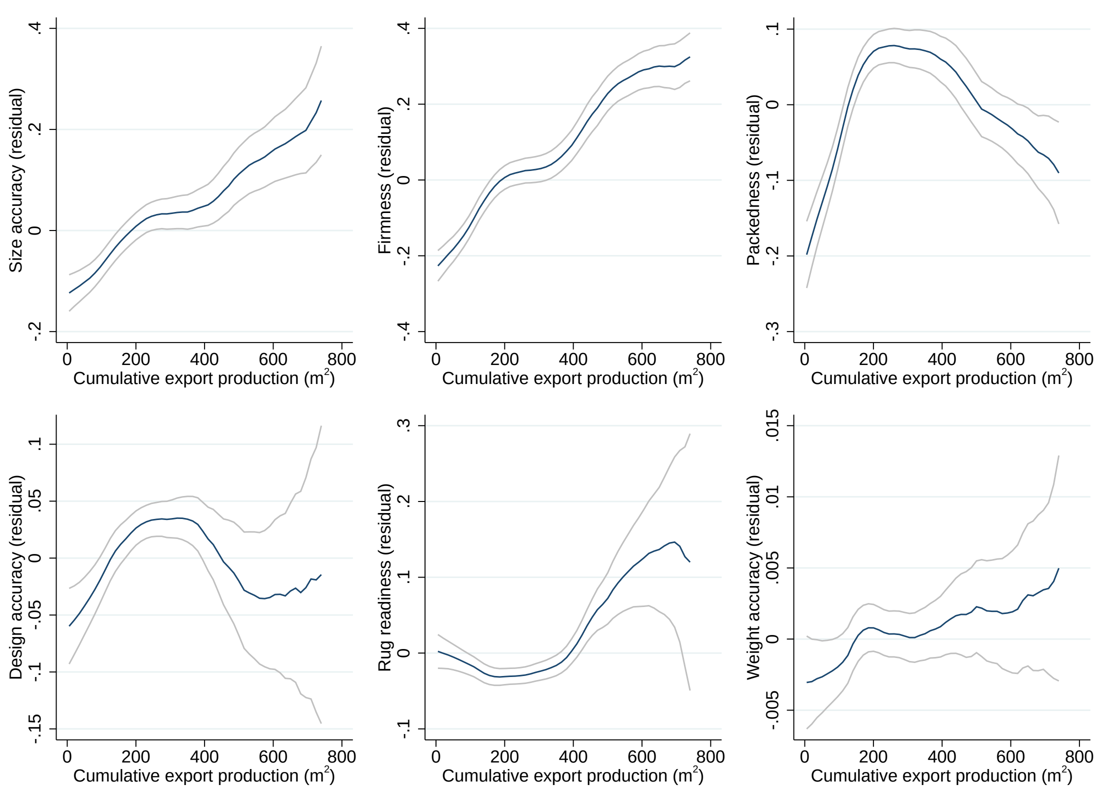


Figure B.4: Quality Problems Noted by Overseas Buyer