Exporting and Firm Performance: Evidence from a Randomized Trial*

David Atkin[†] Amit K. Khandelwal[‡] Adam Osman[§]

First Draft: June 2014
This Draft: April 2016

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 16-26 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.

Keywords: Exports, Quality, Learning-by-Exporting, Productivity, Market Access

JEL Codes: F10, F14, D24

*We thank Abdelrahman Nagy, Magdy El-Mezain, Atef Helmy, Salah El-Gazar, Aslam Ismail, Mabrook Abou Shaheen, Sherif Abou Shady and the firms who participated in our sample. We also thank the staff at Aid to Artisans 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, Christopher Woodruff, and multiple seminar participants for useful comments. We acknowledge generous funding from the International Growth Centre, Private Enterprise Development for Low-Income Countries, Innovations for Poverty Action, Economic Growth Center at Yale University, McMillan Center at Yale University and the Jerome A. Chazen Institute of International Business at Columbia Business School. Yale IRB Approval #1001006247 (Atkin and Osman were at Yale University during the experiment), Columbia IRB Approval #AAAE9678.

[†]MIT, *email*: atkin@mit.edu
[‡]Columbia Business School, *email*: ak2796@columbia.edu
[§]University of Illinois at Urbana-Champaign, *email*: 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 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", and the last two decades have seen a tripling in the number of national export-promotion agencies (Lederman et al., 2010). Central to these programs achieving this goal is the belief 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). In the presence of learning-by-exporting, trade generates efficiency gains which narrow this productivity gap and magnify the gains from trade (e.g., Alvarez et al. (2013)).

Despite the pervasiveness of these initiatives, there is still an ongoing debate as to whether exporting has a causal impact on measures of firm performance. Moreover, if performance does improve, it is unclear whether such improvements occur through learning-by-exporting—outward shifts in the production possibility frontier (PPF)—or simply through movements along the PPF. There are two central challenges to answering these questions. First, more productive firms select into exporting (see the survey by Melitz and Redding 2014). This selection has plagued empirical attempts to identify the causal 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. The literature commonly uses revenue-based total factor productivity (TFPR) measures which also reflect changes in markups, the product mix, and product quality (de Loecker and Goldberg 2014). This is problematic for identifying learning-by-exporting since theoretical work suggests all three are likely to change with exporting.[1] While quantity-based total factor productivity measures (TFPQ) solve problems related to changing markups, standard datasets do not provide the level of detail required to account for changes in product specifications or quality.[2] Hence, if trade causes firms to change along these dimensions, measured improvements in quantity-based productivity measures may simply reflect movements along the PPF, rather than outward shifts of the PPF.

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

[1] See Edmond et al. (2015), Bernard et al. (2011), and Fajgelbaum et al. (2011) for examples of models where trade affects markups, the product mix and product quality, respectively.

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

The random assignment into exporting directly addresses the first of the two challenges detailed above: selection of firms into exporting. We provided a subset of firms with the opportunity to export handmade carpets to high-income markets. To provide this opportunity, we partnered with a 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 sell to high-income markets.

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

Thanks to the randomization procedure, the causal effects of exporting are identified by comparing mean outcomes between treatment and control firms. We find that the opportunity to export raises the overall performance of firms as measured by profits—treatment firms report 16-26 percent higher profits relative to control firms. The substantial increase in profits is perhaps not surprising given that firms were provided with a positive demand shock, but is interesting given the more moderate profit impacts the literature has found when exploring supply-side interventions such as credit access (Banerjee, 2013).

The primary focus of this paper is to understand the mechanisms driving the profit increases. Despite increases in output prices and labor hours, we observe a decline in total output (m^2 of rugs produced) among treatment firms. These findings suggest that buyers from high-income countries demand higher-quality rugs that take longer to produce. Indeed, our quality assessments show that the rugs produced by treatment firms score significantly higher along virtually every quality dimension. At the same time, "unadjusted" productivity measures—those that do not control for rug specifications and quality (e.g., output per hour)—fall by 24-28 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 output per unit input of a firm depends on both rug specifications and an output efficiency parameter χ_a; high-specification rugs take longer to weave and, *ceteris paribus*, firms with higher χ_a produce more output per unit input. Quality also depends on rug specifications and a quality efficiency parameter, χ_q, and is increasing in both. The export opportunity exposes firms to buyers in high-income markets, and these buyers are willing to pay more for quality than domestic buyers. As long as firms find it profitable to do so, they will raise specifications, and hence improve quality. Under this first mechanism, firms already know how to manufacture high-quality rugs and the opportunity to export simply induces a movement *along* the PPF. That is, there is no change in either efficiency parameter.

A second mechanism involves an increase in the efficiency parameters induced by exporting: learning-by-exporting. This learning can come about through transfers of knowledge from buyers to producers, or from learning-by-doing if such learning would not have happened without exporting (a distinction we return to). Learning-by-exporting is an outward shift of the PPF which can occur either by raising χ_a (producing more output per input conditional on specifications) or raising χ_q (producing higher quality conditional on specifications). When these increases in efficiency are biased towards the production of high-quality rugs, both rug quality and profits will rise. Of course, the two mechanisms are not mutually exclusive, but the presence of learning-by-exporting is important because it implies larger gains from trade (Alvarez et al., 2013). Unlike previous studies, the random assignment of export orders and our detailed data collection allow 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 "unadjusted" productivity falls). If firms only moved along the PPF, specification-adjusted quality and productivity would remain constant. Second, at the end of our experiment, we asked all firms 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 specification-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 partly 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 (despite the fact that, even in the domestic market, the knowledge has a return that substantially exceeds the costs of providing it). Taken together, the evidence strongly supports the presence of learning-by-exporting.

As with any industry- or country-specific study, it is important to acknowledge issues relating to external validity. In terms of the context of our study, the firms in our sample are small—typically having only one full-time employee—and production is not automated. Hence, our study has little to say about learning-by-exporting for large firms manufacturing complex products. Of course, it is precisely their small size that allows us to assemble a large sample necessary for inference; and the fact that they manufacture products using the same basic technology improves statistical power (Bloom et al., 2013). The firms also export via an intermediary, rather than directly, but indirect exporting is common for the Egyptian rug industry and in other industries and countries.[3]

In terms of the experiment itself, our treatment induces exporting by reducing matching frictions between firms and sophisticated foreign buyers. Such frictions are of interest both for theoretical and policy reasons. As noted by Allen (2014), matching frictions comprise a sizable component of trade frictions,[4] while Lederman et al. (2010) note that reducing matching frictions for small- and medium-sized firms (SMEs) is a key goal for export promotion agencies. That said, our experiment does not reduce the trade frictions more typically studied in the trade literature such as tariffs or transport costs. While ultimately an empirical question requiring further research, we conjecture that exporting to high-income markets will generate similar learning for developing-country SMEs however they are induced to export (whether by reducing tariffs, trade costs or matching frictions; or by subsidizing matching costs).

Two more caveats are necessary. Given the difficulties we document in generating orders and

[3] World Bank Enterprise Surveys reveal that 36 percent of exporters across countries and sectors use an intermediary (and 62 percent for exporting firms with 5 or fewer employees).

[4] Allen (2014) estimates that such frictions explain half of the overall variation in trade costs. See also the recent survey by Donaldson (2015).

the implicit labor costs of our time, it is unclear if it is efficient for our sample firms to pay the high fixed costs required to find sophisticated foreign buyers. 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 (both questions that would require an entirely different experimental design involving a large number of NGO-led interventions). Instead, the goal is to identify the presence of learning-by-exporting. Finally, given the nature of our experiment, we are unable to distinguish exporting from selling to rich domestic buyers that demand high quality. However, these buyers are scarce in developing countries, and as the literature on quality upgrading we discuss below suggests, the presence of buyers demanding high quality may be the most pronounced difference between internal and external trade for developing-country firms.

Our results relate to a number of papers that span the trade and development literatures. Most directly, we contribute to a voluminous literature that seeks to identify the existence of learning-by-exporting. The evidence from these studies is mixed.[5] 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 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 findings 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 exporting exposes firms in developing countries to sophisticated buyers who demand higher quality.[6] 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.[7] In addition to our randomization methodology, the quality lab and the 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

[5]For example, Clerides et al. (1998) and Bernard and Jensen (1999) conclude that firms self-select into export markets. In contrast, others use different techniques to deal with selection (e.g., matching estimators or instrumental variables) find some support for the learning hypothesis; see de Loecker (2007), Park et al. (2010), Marin and Voigtlander (2013) and de Loecker (2013). Keller (2004), Wagner (2007) and Harrison and Rodriguez-Clare (2010) survey the literature.

[6]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.

[7]Papers that infer quality using structural approaches include Khandelwal (2010), Hallak and Schott (2011), and Feenstra and Romalis (2014).

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 five step approach to detecting learning-by-exporting. Section 7 concludes.

2 Research Setting

2.1 The Industry and the Location

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

ATA's program in Egypt followed their standard protocol for generating successful exporting relationships between small-scale developing-country producers and high-income OECD markets. First, ATA explores the country in question for products that would both appeal to consumers in high-income markets and be priced competitively. Once candidate products are found, ATA identifies a lead intermediary based in the developing country. The lead intermediary assists in finding small-scale producers 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 and draws on its 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. ATA matches the intermediary with foreign buyers and then the intermediary aggregates orders to spread the fixed costs of exporting across many small producers. The ultimate objective is to foster self-sustaining relationships whereby the intermediary maintains or expands its clients without further assistance from ATA.

This process of exporting through intermediaries is not uncommon in other settings.[8] Among manufacturing firms in the World Bank's Enterprise data, 36 percent of exporters use an intermediary with this number rising to 62 percent when we restrict attention to firms with five or fewer

[8] 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. Their Chinese customs data suggest that exporting via intermediaries is particularly common in the rug industry with 52 percent of exports in HS Code 570231 ("Carpets and other textile floor coverings, wool") going through intermediaries compared to 20 percent of overall exports. The need for such intermediation between suppliers and buyers has also been noted by Rauch (1999) and Feenstra and Hanson (2004).

employees to facilitate comparison with our context.

Alongside ATA, we searched for viable Egyptian products and identified handmade carpets from Fowa as having potential. In terms of the industry choice, both the handmade craft industry and the rug industry are large and important sources of employment in Egypt, as well as in many other developing economies.[9] Turning to the location, Fowa is a peri-urban town with a population of 65,000 located two hours southeast of Alexandria.[10] The town is well known for its carpet cluster which contains hundreds of small firms that use wooden foot-treadle looms to manufacture flat-weave rugs—a product in which Egypt has a strong historical reputation. Crucially, we also identified a local firm, Hamis Carpets, to serve as the lead intermediary. Hamis is the largest intermediary in Fowa and accounts for around 20 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.

The firms in Fowa typically consist of a single owner who operates out of a rented space or sometimes his (all producers in our sample are men) home. Family members or hired labor assist with setting up the loom and the finishing stage. The process of producing rugs is standardized across firms. The two key inputs beyond the loom and labor are *warp thread*—wool or cotton thread that spans the length of the rug and is not visible on the final rug but are necessary to hold the rug together and *weft thread*—the colorful threads weaved between the warp threads using a shuttle. Online Appendix C provides additional details regarding the production process.

Firms self-identify as specialists in one of four flat-weave rug types: duble (the focus of this paper), tups, kasaees and goublan.[11] The average duble rug destined for domestic markets is sold by firms for LE42.5 (about \$7 at the prevailing exchange rate) and requires 5.9 hours of labor, per m^2.[12] After accounting for input costs, hourly wages are roughly LE3 (\$0.48).

Within a particular rug type, quality can vary substantially. Broadly speaking, there are two determinants of quality. First, higher quality is associated with more demanding specifications. Specifications are codifiable attributes of the rug that are typically chosen by the buyer; for example, the number of colors, the thread count or the type of input thread. Figure 1 shows one such specification sheet from a foreign buyer in our experiment. Second, higher quality is associated with better weaving technique. For example, how flat the rug lies on a hard surface is determined by how skillfully the warp and weft thread are installed on the loom, and whether the threads are held correctly while weaving. Similarly, how well defined the corners are, how accurately the design was followed, and whether the rug adheres to the desired size specifications depend on weaving skill. These attributes of the rug are both difficult to codify and depend on the skill of the firm.

[9] Online Appendix B provides statistics in support of this claim.

[10] The average income per capita in the governorate containing Fowa is \$3,600, well below the national average of \$6,500 (both PPP-adjusted).

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

[12] 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.

2.2 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 first brought the CEO of Hamis to the US for a training course. Hamis was then provided with marketing support—both by displaying the new products at various international gift fairs and by directly introducing Hamis to foreign importers or retailers through a US-based rug intermediary.[13] Hamis and potential foreign buyers would discuss pricing, delivery time, and product specifications (design, colors, materials, and so forth; recall Figure 1 for an example of how these specifications are codified after these discussions). Hamis would then organize the production of sample orders, either from its in-house weavers or from one of the treatment firms in our sample.[14]

The majority of rugs demanded by foreign buyers are duble rugs, although one client ordered kasaees rugs. There have been no orders for goublan rugs, even though the local market in Egypt perceives these rugs to require the most skilled weaving techniques; the painting-like style of goublan rugs is unlikely to appeal to high-income OECD buyers (see Figure C.1 for an example). Likewise, domestic-style duble rugs did not attract interest from abroad. Instead, it appears that high-income OECD buyers prefer "modern" designs, as illustrated in Figure 2.[15]

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. As of June 2014, buyers continue to place large, regular orders. Figure 3 reports that cumulative export production between December 2010 and June 2014 (the end of our experiment) has 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

We designed the following export-market-access intervention. We drew a sample of small rug-producers (described in more detail in the next section). The firms were divided into two groups, treatment and control. As described above, the local intermediary, Hamis Carpets, secured export

[13] ATA's 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 for several reasons. First, we sponsored the CEO's visit to the New York International Gift Fair in January 2013. Second, we provided a quarter of the capital ($7,000) to finance a sample order for a new client which was ultimately unsuccessful. Third, we provided $500 a month to offset costs of participating in the experiment (conducting rug quality surveys, filling out order books etc.). Finally, the CEO believes that showing how exporting improves the livelihoods of the local population will be good for promoting Fowa's weaving industry.

[14] 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.

[15] One rug style produced by our sample firms retailed for $1,400 in a chain of high-end furniture stores in the United States.

orders with ATA's marketing assistance. Treatment firms were then visited by a representative of Hamis Carpets and provided the opportunity to fill an *initial* export order. More precisely, Hamis Carpets showed them the rug design, explained that the carpet would be exported to 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 by trading off the desire to have a reasonable sized initial order and the need to have enough orders to treat the firms. Hamis was free to choose the price offered to the producers based on the specifications of the rugs (prices we analyze in detail below). To ensure all rug orders were consistent across producers, Hamis provided the input thread and two loom components that are design-specific, the reed and the heddle (see Online Appendix C). At the same time, as is typical in 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 was allowed to continue to place them with 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 firms. In contrast, the control firms were not visited initially, and Hamis Carpets was forbidden from sourcing from them through the duration of the experiment.[16] 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 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

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

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 the intermediary received) that came from international markets.

3.2 Sample Details and Takeup

To obtain the sample of firms, we carried out a recruitment drive in Fowa in July 2011. To be eligible, the firm had to have fewer than 5 employees, work on their own account (meaning that they bought their own inputs), and have never previously worked with Hamis. The recruitment drive generated a sample 303 firms that specialized in one of the four rug types described in Section 2.1.[17] Anticipating that we would not secure orders for every type of rug, we stratified the sample on the type of rug produced and the loom size (which determines the maximum size of rug that can be produced). Within each stratum we randomized firms into treatment and control using a random number generator with strata containing an odd number of firms assigned one additional control firm. For reasons that will become 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.

Securing sufficient export orders to treat every firm in the treatment group with an 110m^2 initial order proved difficult. As detailed in Section 2.2, we were only successful in generating large and sustained orders for one of the four rug types, duble rugs, and even then only one and a half years after the baseline survey. For this reason, we were only able to offer firms the opportunity to produce duble rugs and we could only offer each firm multiple smaller orders of 10-20m^2 spread out over many months instead of the full 110m^2 order in one go. This resulted in very low takeup (16 percent) among specialized producers of the three other rug types with the firms not taking up citing an unwillingness to switch rug types (for a more detailed discussion, see Online Appendix D).[18] Even among the 79 specialist duble producers, 39 firms were treated and only 14 firms took up. The firms not taking up cited an unwillingness to jeopardize their existing dealer relationships for the small 10-20m^2 order sizes we were able to offer at the time.

Once large and sustained export orders for duble rugs arrived in June 2012, it was feasible to offer the full 110m^2 initial order in one go. Given the opportunity to implement the experimental design we originally intended, we asked our surveyors to locate every remaining duble firm in town that satisfied our sample criteria (i.e. less than 5 employees, works on own account, and never worked with Hamis). We found 140 additional duble firms in this manner which we refer

[17] We hired a local NGO to go street by street looking for rug-making firms in homes and workshops. Firms were asked about the number of employees, type of rugs being produced, width of loom and whether they would be willing to be part of a research experiment on exporting in the rug industry. Based on our power calculations, we stopped the recruitment drive when the NGO had identified 300 firms. As there are far more than 300 rug-making firms in Fowa, the first recruitment drive provided us with a sample of firms in central areas of Fowa that were operating and observable at the time of recruitment.

[18] We were able to initially offer the 19 treated kasaees firms orders for kasaees rugs, but we were unable to secure follow-up orders for the 5 firms who took up.

to as “Sample 2”. We stratified on loom size as before, and used a random number generator to select 35 firms for the treatment group (with the number of treated firms determined by Hamis’ capacity constraints). Consistent with the claim that takeup in Sample 1 was low due to the small order sizes, column 5 of Table 1 shows that 32 of the 35 treated Sample 2 firms tookup when offered the full 110m^2 initial order in one go. All of the duble firms in both Sample 1 and Sample 2 that took up the opportunity were “successful” in the sense that they delivered the 110m^2 that constituted the initial order and then received subsequent orders from Hamis.

Given that we only secured large and sustained export orders for duble rugs and that very few non-duble firms were willing to manufacture this rug type, we were essentially unable to treat the non-duble strata and so exclude them from our analysis. 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?

For the analysis in the rest of the paper, we combine the duble strata from Sample 1 with the firms in Sample 2—who are exclusively duble firms—to form the *Joint Sample* of 219 firms (74 in treatment, 47 of which took up). For completeness, Online Appendix F presents the results for the two samples of dubs firms separately,[19] as well as results for the non-duble strata.

3.3 Data

We collected multiple rounds of data at around 4-month intervals. In total, Sample 1 was interviewed seven times and Sample 2 four times. The baseline round occurred before treatment firms were provided with the opportunity to export and included questions on: (a) firm production; (b) rug quality; and (c) household and demographic characteristics.. We administered an abbreviated survey on firm production and rug quality in each follow-up round, except for follow-up round 5 for Sample 1 and follow-up round 3 for Sample 2 where we again administered the full survey. The survey timeline is shown in Table 2.[20]

The production module records production activity for the month preceding the survey interview. All nominal variables are converted to real values using the official Egyptian CPI. We collect measures of profits, revenues, expenses, output quantity and prices, input quantity and prices, total labor hours worked, and the specifications of the rugs produced that month. These specifications include: (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 weft thread used per m^2 of the rug

[19]Other than the lower takeup, results are similar across the two samples: 17 out of 18 treatment-on-the-treated coefficients are not significantly different when comparing across samples.

[20]We hired an Egyptian survey company to conduct the baseline survey on Sample 1. The company trained a new enumerator who was responsible for the first follow-up on Sample 1. Unfortunately, we discovered that this enumerator did not actually interview all of the firms and entered in fake data for some, so we discard this round. We immediately fired the enumerator and hired new employees in January 2012 who conducted all subsequent surveys. We managed these employees directly and implemented a stricter auditing procedure as well as back checks using external evaluators to ensure data integrity.

(thread count); (4) the type of weft thread used (e.g., Egyptian wool, cotton, etc.); (5) the number of colors used in the rug; and (6) 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:[21] (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.[22] 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 2.1, 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, and weight accuracy.

We collected a third set of quality and productivity measures in June 2014 by asking firms to manufacture an identical domestic-specification rug using identical inputs and a common loom in a work space we had rented (our quality lab). Production was timed and 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.

3.4 Summary Statistics

Table 3 shows baseline balance between the treatment and control groups. The table reports regressions of each variable on a treatment dummy and strata fixed effects, and reports the constant (the mean of the control firms) and treatment coefficient (the difference between control and treatment means). Panel A shows summary statistics for the household characteristics of the firm owner. The mean age in the control group is around 51 years and, on average, firms have slightly more than 37 years of experience working in the rug industry. Roughly 63 percent of firm owners are illiterate. The average household size is 4.2.

Panel B reports statistics from the rug business. Monthly profits from the rug business averages LE664 ($105). Firms report 247 labor hours in the previous month, which amounts to around 22 days of work at 11 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 50m^2 and only about 12 percent of firms have ever knowingly produced rugs for the

[21] The Sample 1 baseline survey recorded 6 quality metrics to which we subsequently added 5 more metrics.

[22] 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.

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: treatment firms report *lower* quality scores at baseline. The final row of Table 3 reports attrition across survey rounds. Attrition has been relatively low with a non-response rate of approximately 11 percent per round 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 use a straightforward specification to assess the impact of the export-market access on firm profits:

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

where y_{it} is the profit measure, $Treatment_i$ is an indicator variable that takes the value 1 if firm i is in the treatment group, τ_t are time period fixed effects, δ_s are strata fixed effects and y_{i0} is the value of the dependent variable at baseline.[23] Since (1) controls for the baseline value of the dependent variable, we cannot include observations from the baseline survey in the regression.[24] Since 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 the treatment effect to take account of the fact that not everyone was actually treated:[25]

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

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 55 percentage points from a baseline of 13 percent. We also report the TOT specification, which

[23] We note that any effects we find will be attenuated if there were spillovers to control firms. However, we find no support for geographic spillovers between treatment and control firms (see Table E.3).

[24] Alternatively we could use all survey rounds, include firm fixed effects, and interact $Treatment_i$ with a post-baseline dummy. We prefer a specification with baseline controls because if the dependent variable is measured with noise and not strongly autocorrelated, as is the case for business profits (an autocorrelation is 0.33 among control firms in our data), the fixed effects estimator will perform more poorly in an experimental study than the ANCOVA estimator in equation (1); see the discussion in McKenzie (2012). For comparison we also report the key results using firm fixed effects in Table E.4.

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

suggests even more dramatic increases.[26]

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 survey questions that ask firms to report their total revenues and total costs from the previous month. The third measure constructs profits from the production modules that contain detailed information on prices and quantities of inputs and outputs. The 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 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 measure 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

The top panel of Table 5 shows the results of running specifications (1) and (2) on logged values of the four profit metrics. 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 monthly profit measure. The ITT coefficient is 0.26, implying that the export treatment increases monthly profits by approximately 26 percent. The TOT coefficient is, not surprisingly, larger at 42 percent and also statistically significant. 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 21 and 37 percent, respectively. The reason these point estimates are similar to columns 1A and 1B may be because the firms in our sample typically

[26] Note that the ITT and TOT do not perfectly 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.

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 19 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 37 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 baseline and post-baseline survey rounds (regressing profits on a post-baseline dummy for control firms yields a coefficient of 0.041 (s.e. of 0.055).

These regressions indicate that the export treatment causally increases profits by between 19-37 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, the lower panel of Table 5 examines 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 20 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 the panels suggests that total owner hours increased by 5 percent.[27] However, the basic message remains the same: the opportunity to export raised profits per owner hour by 16-25 percent.[28] In the next section, we examine this increase in hours in more detail.

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

[27] 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 across the two panels does not match the increase in total hours inferred from the other columns.

[28] Of course this will not correspond to true profits if owners of treatment firms increased their effort per hour. We were unable to devise survey questions that could accurately capture worker effort, but we did ask about job stress measured on a scale of 1 to 5. Reassuringly, when we regress job stress on log output per hour, treatment and the interaction of the two (plus round and strata fixed effects), the coefficients are not individually or jointly significant (F-stat of 1.40).

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:

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

where p is the price a firm receives for one unit of rug. The quantity of rugs produced is x, w is the wage paid for each hour of labor l, r is the rental rate on capital k, and F is a fixed cost of production. We do not include input costs in (3) since a large majority of firms (74 percent) receive raw material inputs from their intermediary and hence do not pay for these expenses. Hamis follows this industry norm. For the subset 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.

Panel A of 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 43 percent increase in prices with the opportunity to export while the TOT indicates a 78 percent increase. Thus, part of the profit increase from exporting is coming from significantly higher prices per m^2 of rug for export orders.

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 -26 percent while the TOT is -47 percent; this is a large *decline* in output among treatment relative to control firms.[29]

Columns 3 and 4 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 and the log number of active looms in use k. The ITT estimate indicates a labor increase of 5 percent and the TOT is 8 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. Consistent with the lack of extensive margin effect, there is no significant change in the number of looms.

Finally, we turn to fixed costs F in columns 5A and 5B. The proxy we use is the size of the warp

[29]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.

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 15 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.

Panel B of Table 6 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. Note 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 20 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 both material type (cotton, wool, polyester, silk or various blends), thickness and material grade (e.g., Egyptian wool or polyester). 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 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 provides 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 Unadjusted Productivity Measures

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 7 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 7 restricts the coefficients on the treatment dummy to be identical across the various quality metrics (recall they were all run in a single stacked regression).[30] Given the previous results, it is not surprising that we obtain positive and statistically significant ITT and TOT estimates when we do this. On average, quality (on a scale of 1 to 5) is 0.79 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 two productivity measures that do not adjust for changes in product specifications or quality: unadjusted output per labor hour and unadjusted total factor productivity (TFP). Unadjusted output per labor hour comes from firms' responses to the question: "how long does it take you to make 1 meter squared?".[31] This measure abstracts from capital inputs because there is very little variation in capital across firms: 92 percent of firms use one loom and no firm in our sample purchased (or rented) an additional loom during the period of our study.[32] The second measure, unadjusted TFP, is the residual from the estimation of a production function with both labor and capital (see Appendix A for further details). Panel A of Table 8 shows that unadjusted output per hour falls 24 percent among treatment relative control for the ITT specification, with even larger TOT effects. Panel B presents the unadjusted TFP measure and we find a similar 28 percent decline for the ITT specification.

5.4 Mechanisms

The finding that quality rises and unadjusted productivity 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 re-

[30] 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).

[31] Another way to measure output per hour is to divide total output by total hours worked in the month. Related to the discussion of the profit measures above, we believe the direct measure is less noisy. We find virtually identical results using the alternative measure (available on request).

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

turns 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 (the full cost structure is broken down in Table E.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 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

6.1 A Framework for 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:

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

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

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

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

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 $b > 0$.[33] Rug output and quality are determined by two production functions, both of which depend on efficiency parameters that depend on the skill of the firm as well as a choice variable: the product specifications of the rug indexed by λ. (Recall that specifications are codifiable rug attributes that buyers agree upon before ordering; see Figure 1 for an example of such an agreement.) High-λ rugs have more demanding specifications,

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

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

The production function for output $x(\lambda, l, k)$ has two components. Labor and capital inputs are mapped to output through $f(l, k)$ and output per unit input is determined by the function $a(\lambda; \chi_a)$, a TFP metric that is "unadjusted" for rug specifications.[34]

Unadjusted TFP $a(\cdot)$ is necessarily decreasing in λ since rugs with more demanding specifications require more inputs. The function $a(\cdot)$ is also increasing in χ_a, an efficiency parameter which governs how quickly a firm produces rugs of a particular set of specifications with a given set of inputs. Collecting these two derivatives:

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

Quality is determined by the function $q(\lambda; \chi_q)$ which we assume is increasing in product specifications as quality is achieved in part through more demanding specifications. Additionally, quality increases in χ_q, an efficiency parameter 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 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 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 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 unadjusted TFP a alongside improvements in quality q through either mechanism, we rearrange the total derivatives of the first order conditions with respect to λ and l. As long as there are diminishing marginal returns to raising specifications (i.e, concavity of q and a in λ), firms choose to raise specifications with an increase in the price of quality b; hence, equilibrium quality q^* rises and equilibrium unadjusted TFP a^* falls:[35]

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

Similarly, specifications rise with an increase in the quality efficiency parameter χ_q and hence q^*

[34] We abstract from material inputs in the production function since, as discussed earlier, the intermediary provided raw materials to the firm.

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

rises and a^* falls,

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

as long as the complementarity between λ and χ_q in producing quality is sufficiently large—i.e., weaving skill is particularly valuable for the production of high specification rugs, $\frac{\partial^2 q}{\partial\lambda\partial\chi_q} > 0$ (for example, higher thread-counts require more dexterity to weave).[36] Intuitively, the strong complementarity ensures that it is profitable for the firm to raise specifications since quality, and hence price per m^2, increases faster than a declines. The same pattern can arise with an increase in the output efficiency parameter, χ_a, as long as there is a sufficiently large complementarity between λ and χ_a in producing output—$\frac{\partial^2 a}{\partial\lambda\partial\chi_a} > 0$; for example, skill matters more for speed when producing less-familiar high-specification rugs—to ensure that the slowdown from raising specifications exceeds the direct increase in speed from higher χ_a.[37]

In these cases, increases in b, χ_q or χ_a lead firms to raise product specifications and hence produce higher quality products that sell for higher prices. However, unadjusted TFP falls because high-λ rugs require more 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.[38] 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 effi-

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

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

[38]It is for this reason that the framework excludes marketing capital as a factor in the production function. We believe, consistent with the learning-by-exporting literature, that the knowledge generated from matching with foreign buyers is not anticipated by the firms and hence should be reflected in productivity. See Section Appendix A for a more detailed discussion.

ciency through residual-based TFP. If prices are higher in export markets, TFP measures that do not adjust for prices (which is rarely the case) may suggest learning-by-exporting when firms are just moving along the PPF or obtain a higher markup.[39] In the few cases where price adjustments are made, measuring quantity-based TFP requires comparing products with identical specifications and quality-levels. This is typically achieved by focusing on homogenous goods like concrete and block ice (e.g, Foster et al., 2008), or using product dummies based on administrative classifications and inferring quality from prices and market shares (e.g., de Loecker et al., 2014). In contrast, we exploit our rich panel data and experimental variation to solve these measurement issues.

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

1. In Step 1, we use our detailed data on product specifications to show that although our unadjusted productivity measures—corresponding to $a(\cdot)$ above—fall with the opportunity to export (recall Table 8), specification-adjusted productivity *rises*—consistent with χ_a increasing. We also show that specification-adjusted quality rises consistent with χ_q increasing. If there is no learning-by-exporting, specification-adjusted productivity and quality should be *unchanged* as there is no change in the efficiency parameters.

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 productivity evolve over time as cumulative export production increases, consistent with a learning process. In contrast, if firms simply moved along the PPF we would expect a discontinuous jump upon exporting as firms immediately move to new quality and productivity levels.

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, and do not pay, even implicitly, for the knowledge they receive from the intermediary.

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

6.2 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 now repeat the quality and productivity regressions above but control in various ways for the specifications of the rug being manufactured at the time of the survey visit. Recall from Section 3.3 that we have six dimensions of rug specification. Although imperfect, we note that many studies simply control for product differences through product fixed effects based on statistical classifications. Our first specification—the type of rug—is the analogous control, although it uses a much finer classification than standard trade classifications (e.g., all of our seven rug types would fall within the U.S. HS ten-digit classification 5702311000). The remaining specifications, such as thread count or design difficulty, are rarely observed by researchers. Such controls are possible because there is overlap in rug specifications across firms selling to domestic and foreign markets. This overlap can be seen in Figure 4 which plots the distribution of each of the six specifications separately for firms that are producing rugs for export (i.e. $Takeup_{it} = 1$) and those that are not. Note that if our characteristic controls are very crude, that will tend to bias our findings towards the unconditional results we found in Tables 7 and 8. Hence, the prediction that productivity should *rise* conditional on specifications is particularly informative since unadjusted productivity *fell*.

We present three sets of results. The first approach regresses stacked quality, output per hour and unadjusted TFP on treatment (or treatment instrumented with takeup) as in Section 5.3 but now also includes controls for the six specifications.[40] The second approach goes further towards ensuring that the treatment and control firms we compare are making identical rugs by including fixed effects for each of the 435 unique combinations of the six specifications. Since approximately one third of firm-round pairs are making unique rugs, the cost of this approach is that we lose a significant number of observations. These two approaches explore differences in unadjusted productivity and quality between treatment and control conditional on rug specifications. The third approach follows from equation (5) which suggests that we can directly infer χ_a from the residual of a production function estimation that includes specifications. As discussed in more detail in Appendix A, we use control firms to estimate this production function and then calculate specification-adjusted TFP for each firm and round using the estimated coefficients on labor, capital and specifications. We construct similar measures for quality and output per hour by regressing these variables on rug specifications in the control group and use the resulting coefficients to construct adjusted metrics (actual minus predicted).[41] This procedure also partially

[40] Since the regressions also include controls for the baseline values of the dependent variable, we also include baseline values of the specifications in the controls.

[41] For the baseline of Sample 1, we did not record the market segment or the rug difficulty. We replace these missing

addresses a second issue. Although treatment is exogenous by design, the specification controls in the first two approaches may be endogenous. Of course, if higher-ability firms selected into higher-specification rugs in the control group, the coefficients on rug specifications will be biased. However, if anything this bias will lead us to find no productivity gain.[42]

Panel A of Table 9 reports the approach with specification controls, Panel B those with fixed effects for specification combinations, and Panel C those with specification-adjusted dependent variables. Before discussing the effects on quality and productivity, it is reassuring to note that the specification controls in Panel A have the signs we assumed in the model: More difficult rugs are associated with higher quality and lower unadjusted productivity, while those destined for lower segments of the market are associated with lower quality and higher unadjusted productivity. And, the R-squared rises substantially (increasing from 0.39 to 0.64 in the quality ITT, and from 0.18 and 0.26 to 0.57 and 0.62 in the productivity ITTs) suggesting that the rug specifications have substantial explanatory power.

Turning to the treatment effects in Table 9, recall that without conditioning on specifications (i.e. Panel B of Table 7 and Panel A of Table 8) quality rises and the productivity measures fall. Conditioning on specifications using the three approaches described above, quality again rises significantly but the signs on the productivity measures flip from negative to positive. (The ITT productivity coefficients correspond to productivity increases between 14 and 31 percent, all significantly different from zero.) That is, conditional on making similar rugs, treatment firms are making them faster than control firms. If firms are just moving along the PPF, treatment and control firms should not differ in their productivity after conditioning on product specifications. However, the results suggest a rise in the efficiency parameters χ_a and χ_q.

6.3 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 data 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.[43] We chose rug specifications that mimicked a popular rug design sold at mid-tier domestic retail outlets in Egypt (the specific rug is shown in Figure 2). 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).

values with the corresponding values from the subsequent survey round.

[42]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 would likely be biased downward (because lower-ability firms in the treatment group would "appear" less productive if the coefficients used for the adjustment were biased).

[43]The owner was paid a market wage plus compensation for having to make the rug in an external location. See Online Appendix D.3 for further details.

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 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. We provide the experimental protocol we used for this Step in Online Appendix D.3.

In Panel A of Table 10, we report results separately for each quality metric.[44] Quality is significantly higher among treatment firms for all 9 quality dimensions.[45] Reassuringly, treatment firms also score higher along every dimension using the professor's quality assessments.

Panel B of Table 10 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 0.64. 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 three quarters of a standard deviation above control firms.

Panel C of Table 10 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. Although not significant, treatment firms took less time not more. That is, despite manufacturing rugs with higher quality metrics, treatment firms did not spend more time weaving.

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. It also seems unlikely that treatment firms put more effort into weaving the rug because they were worried poor performance would jeopardize their relationship with Hamis. Firms were not informed of any link between the quality lab and Hamis Carpets—we hired a new staff member to run the lab and told firms that the order was from a buyer in Cairo—and if firms did believe there was a link it is just as plausible that control firms put in extra effort to impress

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

[45] 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.

Hamis in order to gain export orders.

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

6.4 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 6.6 for a discussion of adjustment costs), learning processes typically take time. Hence, we write the two efficiency parameters in period t as:

$$\chi_{k,t} = h_k \left(\sum_{t'=0}^{t} \left(x_{t'} \mathbf{1} \left[\text{export}_{t'} \right] \right) \right) \text{ for } k = q, a. \tag{10}$$

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 quality or productivity measures on firm fixed effects as well as round fixed effects.[46] 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). Figure E.1 presents similar plots using the partially linear panel data estimator proposed by Baltagi and Li (2002).

Figure 5 shows these residual plots for the productivity measures as well as the stacked quality measure.[47] The upper left graph reports the unadjusted output per hour measure; the figure indicates a decline in productivity until about 400m^2 after which it starts to rise. We draw similar

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

[47] Analyzing the Joint Sample is complicated here by the different timelines across samples. Figure E.4 shows the cdf of total export production across the two samples and Figure E.5 plots days since first order against export production for each firm. As Sample 1 firms started exporting earlier they drive all the variation at high values of cumulative exports but not at lower values, leading to a spurious discontinuity around this transition. Accordingly, Figure 5 restricts attention to the common support; the range of cumulative exports achieved by Sample 2 firms (0-727m^2). Figures F.1 and F.2 present the two samples separately across their full ranges.

conclusions from unadjusted TFP (middle-left figure). The second column uses the specification-adjusted measures described in Section 6.2. Both specification-adjusted output per hour and TFP rise with cumulative exports, consistent with the initial dip in unadjusted productivity being driven by the move to more difficult product specifications demanded by foreign buyers. The bottom row of Figure 5 presents the analogous learning curve for the stacked quality measures. 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 five months. In the right panel, we observe a similar path when using specification-adjusted quality. The figure suggests much faster learning about quality efficiency $\chi_{q,t}$ than about output efficiency $\chi_{a,t}$.[48]

In contrast to the ITT results in Steps 1 and 2, the fact we allowed Hamis to allocate follow-up orders means that more orders may have been given to firms whose quality or productivity was improving. However, with that caveat in mind, we see the evolution of productivity and/or quality among takeup firms as suggestive of a learning explanation. Figures E.2 and E.3 complement these results by showing similar patterns using four higher-frequency quality metrics collected by the intermediary for each batch of rugs delivered by each firm.

6.5 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:

> We have a problem with our client. As you remember, this client asked for two carpets

[48] 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. See Table E.5 for a parametric version of these results that interacts the treatment indicator with round dummies and finds learning concentrated in the first follow-up round.

with fringes in the colour uni 2 and 3. Now after one and a half year using the carpets, the fringes crumble away, as you see on the pictures [reproduced in Figure E.6]. They will have two new pieces and give the whole problem to an lawyer. What to do? *(sic)*

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 11 times. A visit lasted 28 minutes on average. They talked about issues related to design accuracy, the weight of the rug and the tightness of the warp thread on at least half of the occasions. During a visit, the intermediary discussed production techniques to achieve higher quality along these dimensions; firms reported that 91.1 percent of discussions about a particular dimension involved the intermediary providing "information on techniques to improve quality" (as opposed to only pointing out flaws). Table E.6 presents more detailed summary statistics from 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:

$$Quality_{id} = \alpha_3 + \beta_3 Takeup_i \times \mathbf{1}[Talked\_About\_Dimension]_{id} + \gamma_3 Quality_{id0} + \delta_i + \delta_d + Takeup_i \times \delta_d + \varepsilon_{id}. \tag{11}$$

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 (both alone and interacted with $Takeup_i$) 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 11 reports the results. We find a positive and statistically significant association between changes in quality and whether the intermediary discussed that quality dimension with the firm whether or not we use our standard quality metrics or our specification-adjusted ones (because of the firm fixed effects and the cross-sectional variation, we cannot include specification controls or specification combination fixed effects). 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 91.1 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]

6.6 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 other steps. Treatment firms score higher quality metrics and have higher productivity once we adjust for rug specifications and produce at a higher quality when they make identical domestic rugs (Steps 1 and 2). 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.

[51] We find a coefficient (standard error) on the interaction of 0.05 (0.04). 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.

Our data allow us to dismiss a simple investments hypothesis. First, we regularly surveyed firms about investments or costs incurred throughout the study. There is no quantitative (or qualitative) support indicating that treatment firms undertook any such investments. For example, no firm reports investing in a new loom or paying to repair existing looms over the duration of the sample. Additionally, we asked treatment firms about the extent to which they practiced weaving techniques, and none report ever practicing techniques. We also find no evidence that, relative to control firms, treatment firms increased scale by hiring more workers, or raised human capital by changing the composition of workers (either paid workers or family members).

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).

A final investment hypothesis is that firms may be investing their time in raising their efficiency parameters by a process of purposeful learning-by-doing (i.e. trading off slower production now for higher returns from exporting once they can produce high-quality). Inconsistent with this hypothesis is the fact that the benefits of the knowledge transfer, even in the domestic market, exceed the time costs that would be incurred if higher quality could simply be learned. To determine the value of improving quality for a firm that sells to the domestic market, we regress profits per hour for non-takeup firms (i.e. those selling to the domestic market) on our quality metrics.[54] Combining these estimates of the domestic returns to quality with the treatment effects estimated in the quality lab of Section 6.3, the benefit for control firms to move to the quality levels achieved by treatment firms would be 9.96 percent higher profits on the domestic market. We assume that it takes 5 months to learn how to produce high quality (the approximate time taken to weave the 200m^2 after which learning stops in the bottom panels of Figure 5). Even if firms do not receive the higher prices associated with increased quality during this learning period, profits would fall by only 10 percent (taking the productivity drop between 0 and 200m^2 in the first panel of Figure 5) and the firm would recover the investment in quality upgrading after 11 months (using an annualized discount rate of 10 percent). Also note that, using the wage the intermediary pays their employees who visit the firms and assuming all the time spent on firm visits is spent discussing techniques, the cost of providing this training is only LE103 ($16) compared to a lifetime net present value for firms of LE10,070 ($1,596). Hence, these calculations suggest that if firms were choosing whether or not to invest in learning the skills needed for high-skill production,

[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 negative and insignificant coefficient of -0.06 (s.e. 0.06).

[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 1.3 percent level. The regression is reported in Table E.7.

they should have already chosen to make this investment to produce for the domestic market.

7 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, the detailed survey collection and our quality lab allow us to make causal inferences about the impact of exporting and to identify the mechanisms through which improvements occur.

We find that profits for treatment firms increase 16-26 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.

ALLEN, T. (2014): "Information Frictions in Trade," *Econometrica*, 82, 2041–2083.

ALVAREZ, F. E., F. J. BUERA, AND J. ROBERT E. LUCAS (2013): "Idea Flows, Economic Growth, and Trade," Working Paper 19667, National Bureau of Economic Research.

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.

BERNARD, A. B., S. J. REDDING, AND P. K. SCHOTT (2011): "Multiproduct Firms and Trade Liberalization," *The Quarterly Journal of Economics*, 126, 1271–1318.

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.

DONALDSON, D. (2015): "The Gains from Market Integration," *Annual Review of Economics*, 7, 619–647.

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

EDMOND, C., V. MIDRIGAN, AND D. Y. XU (2015): "Competition, Markups, and the Gains from International Trade," *American Economic Review*, 105, 3183–3221.

FAJGELBAUM, P., G. M. GROSSMAN, AND E. HELPMAN (2011): "Income Distribution, Product Quality, and International Trade," *Journal of Political Economy*, 119, 721 – 765.

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.

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.

LEDERMAN, D., M. OLARREAGA, AND L. PAYTON (2010): "Export promotion agencies: Do they work?" *Journal of Development Economics*, 91, 257 – 265.

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. (2012): "Beyond baseline and follow-up: The case for more T in experiments," *Journal of Development Economics*, 99, 210–221.

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 Total Factor Productivity

This appendix explains how we measure unadjusted and specification-adjusted TFP.

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). We avoid these measurement issues because we observe output quantities. Moreover, since all firms produce a single product—handmade rugs—issues that arise with multi-product firms and the divisibility of inputs are not relevant in this setting (de Loecker et al, 2014).

The first production function estimation does not control for rug specifications and hence provides our *unadjusted TFP* estimates. We estimate

$$\ln x_{it} = \lambda + \alpha_l \ln l_{it} + \alpha_k \ln k_{it} + a_{it}^{unadjusted} + \epsilon_{it} \tag{A.1}$$

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 $a_{it}^{unadjusted}$ is the firm's unadjusted TFP. We note that there is little variation in the number of looms across firms (92 percent of firms report having one loom), but nevertheless we allow the production function to depend on capital. The error term captures unanticipated shocks as well as an omitted variable, the specifications of the current rug in production.

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

$$\ln x_{it} = \lambda + \gamma_l \ln l_{it} + \gamma_k \ln k_{it} + \mathbf{Z}_{it}'\Gamma + a_{it}^{adjusted} + \nu_{it} \tag{A.2}$$

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

The production functions exclude the real costs incurred to find foreign buyers. Common to the learning-by-exporting literature, we view any productivity gains from exporting as unanticipated by the firm. The evidence in Section 6 supports this claim by showing that productivity improvements come at least in part from knowledge transfers that the firms do not pay for. If the knowledge that comes with exporting were anticipated, one could make the case for including "export marketing capital" in the production function to account for the fact that matches with foreign buyers produce both knowledge and sales. Note that even for management practices, where the case for the gains being anticipated is much more compelling, Bloom et al. (2015) suggest modeling productivity as a function of management practices rather than putting the resources incurred to adopt better management practices in the production function itself.

Although quantity data deal with measurement concerns, there is still potential simultaneity bias since TFP is observed by the firm but not us. We follow the control function approach (Olley and Pakes, 1996) and assume capital is subject to adjustment costs, labor is a flexible input, and we use warp thread quantity as the proxy. We exploit the experimental design by estimating the production function using only control firms, avoiding the need for assumptions on how treatment changes the TFP process.[1] We estimate the production functions using the one-step approach pro-

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

posed by Wooldridge (2009), with l_{it-1} as the instrument for l_{it}, and cluster standard errors by firm. For production function (A.1), we obtain α_l =0.74 (s.e. of 0.53) and α_k =0.30 (s.e. of 0.20).[2] For production function (A.2), we obtain α_l =0.96 (s.e. of 0.46) and α_k =0.20 (s.e. of 0.11). We cannot reject the null of constant returns to scale in either case.

Having estimated the coefficients, we compute $a_{it}^{unadjusted} = \ln x_{it} - \hat{\alpha}_l l_{it} - \hat{\alpha}_k k_{it}$ and $a_{it}^{adjusted} = \ln x_{it} - \hat{\gamma}_l l_{it} - \hat{\gamma}_k k_{it} - \mathbf{Z}_{it}^{'}\hat{\Gamma}$, where $\hat{}$ denotes estimated parameters. Note that the methodology assumes that the parameters of the production function are the same for treatment and control firms. This is a reasonable assumption because all firms produce rugs using identical technology that did not change over the sample period. In fact, since all firms produce a narrowly defined product, the assumption is weaker than existing work that typically assumes identical parameters for all firms within a 2- or 4-digit industry classification.

Appendix References

BLOOM, N., R. SADUN, AND J. VAN REENEN (2015). "Management as a Technology?", Stanford University, mimeo.

DE LOECKER, J. (2013). "Detecting Learning by Exporting", American Economic Journal: Microeconomics, 5(3), 1-21.

DE LOECKER, J. AND P. GOLDBERG (2014). "Firm Performance in a Global Market", The Annual Review of Economics, 6, 201-227.

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

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

[2]For comparison, the OLS of (A.1) gives α_l =0.82 (s.e. of 0.12) and α_k = 0.33 (s.e. of 0.11).

Table 1: Firm Sample and Takeup Statistics

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

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 6th row indicates average output conditional on takeup. The last row reports the standard deviation of output in the first year, conditional on takeup.

Table 2: Survey Timeline

Survey Timeline	Sample 1	Sample 2
Baseline Round 0	*July-Aug 2011	*Feb-Mar 2013
Follow Up Round 1	§Nov-Dec 2011	May-June 2013
Follow Up Round 2	April-May 2012	Nov-Dec 2013
Follow Up Round 3	Sept-Dec 2012	*May-June 2014
Follow Up Round 4	Mar-Apr 2013	
Follow Up Round 5	*July-Oct 2013	
Follow Up Round 6	Jan-Mar 2014	
Quality Lab	June 2014	June 2014

Notes: Table reports the timeline for the data collection by sample. *Supplementary questions about household and firm outcomes are included in both samples Baseline Round 0 surveys and Follow Up Round 5 for Sample 1 and Follow Up Round 3 for Sample 2. §Data from Round 1 for Sample 1 was deemed unreliable and has been discarded for the analysis.

Table 3: Baseline Balance

	Control Group Mean	Difference in Treatment		N
Panel A: Household Characteristics				
Age	51.0	0.9		218
	(0.7)	(1.6)		
Number of years in rug business	37.7	0.2		213
	(0.8)	(1.7)		
Illiterate?	0.63	0.10		214
	(0.03)	(0.07)		
Household size	4.2	0.0		219
	(0.1)	(0.2)		
Digit Span Recall	5.8	0.2		204
	(0.1)	(0.2)		
Cognitive Quiz	0.07	0.03		219
	(0.02)	(0.06)		
Panel B: Firm Characteristics				
Price per square meter	29.1	4.1		218
	(2.4)	(5.6)		
Direct monthly profits from rug business	664	6.3		218
	(36.7)	(69.8)		
Hours worked last month	247	-1.7		218
	(5.6)	(11.7)		
Number of employees	1.09	0.0		218
	(0.0)	(0.1)		
Total produced last month (m^2)	50.0	3.3		218
	(4.3)	(10.0)		
Ever exported?	0.12	0.02		219
	(0.02)	(0.05)		
Average Quality	2.63	-0.13	***	218
	(0.03)	(0.05)		
Joint F-test		1.15		
Attrition in Follow Up Surveys	0.12	0.01		1034
	(0.01)	(0.02)		
Attrition in Quality Lab	0.14	0.02		219
	(0.03)	(0.05)		

Notes: Table presents baseline balance. Each row is a regression of the variable on a constant, treatment dummy and strata fixed effects. The 3rd to last row reports the F-test for a test of joint significance of the baseline variables. Real constructed profits and prices 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 rows report average attrition across all survey rounds and in the quality lab respectively. Significance * .10; ** .05; *** .01.

Table 4: Impact of Intervention on Firms Knowingly Exporting

	ITT (1)	TOT (2)
Indicator for Ever Exported	0.55 ***	0.76 ***
	(0.06)	(0.07)
R-squared	0.33	0.45
Control Group Mean	0.20	0.20
Observations	191	191

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

	Log (Direct Profits)		Log (Reported Revenues - Reported Costs)		Log (Constructed Revenues - Constructed Costs)		Log (Hypothetical Profits)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.26 ***	0.42 ***	0.21 ***	0.37 ***	0.19 ***	0.34 ***	0.37 ***	0.68 ***
	(.05)	(.08)	(.06)	(.10)	(.06)	(.10)	(.11)	(.19)
R-squared	0.21	0.22	0.16	0.18	0.16	0.18	0.19	0.19
Control Group Mean	929	929	931	931	951	951	541	541
Observations	573	573	644	644	685	685	687	687

Panel B: Profit per Owner Hour

	Log (Direct Profits)		Log (Reported Revenues - Reported Costs)		Log (Constructed Revenues - Constructed Costs)		Log (Hypothetical Profits)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.20 ***	0.32 ***	0.17 ***	0.29 ***	0.16 ***	0.28 ***	0.25 ***	0.46 ***
	(.05)	(.08)	(.05)	(.09)	(.05)	(.09)	(.07)	(.12)
R-squared	0.14	0.14	0.12	0.13	0.13	0.13	0.19	0.18
Control Group Mean	3.53	3.53	3.54	3.54	3.55	3.55	5.56	5.56
Observations	573	573	637	637	684	684	687	687

Notes: Table reports treatment effects on different real profit measures, all measured in logs. See text for details regarding each measure. Panel A uses profits as the dependent variable and Panel B uses profits per hour as the dependent variable. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Control group means are reported in Egyptian pounds (LE) in Panel A and LE/hour in Panel B. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 6: Sources of Changes to Firm Profits

Panel A: Components of Profits

	Output Price (LE/m^2)		Output (m^2)		Hours Worked		Number of Looms		Warp Thread Ball (kg)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)	(5A)	(5B)
Treatment	0.43 ***	0.78 ***	-0.26 ***	-0.47 ***	0.05 **	0.08 **	-0.02	-0.04	0.15 ***	0.25 ***
	(.10)	(.19)	(.09)	(.17)	(.02)	(.04)	(.04)	(.06)	(.05)	(.08)
R-squared	0.16	0.15	0.24	0.22	0.12	0.13	0.13	0.13	0.24	0.24
Control Group Mean	28.2	28.2	64.1	64.1	269.0	269.0	1.1	1.1	6.0	6.0
Observations	691	691	676	676	678	678	694	694	600	600

Panel B: Inputs

	Weft Thread Price		Warp Thread Price		Weft Thread Qty (g)		Warp Thread Qty (g)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.20 ***	0.33 ***	-0.04	-0.07	-0.19 **	-0.34 **	0.04	0.06
	(.06)	(.10)	(.03)	(.06)	(.10)	(.17)	(.11)	(.20)
R-squared	0.22	0.24	0.27	0.27	0.23	0.22	0.10	0.11
Control Group Mean	12.8	12.8	18.1	18.1	110.0	110.0	17.8	17.8
Observations	564	564	685	685	677	677	686	686

Notes: Table reports treatment effects on real input price and input quantities, all measured in logs. The TOT regression instruments takeup with treatment. Hours worked are calculated using average daily hours and number of days worked last month. 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: Impact of Exporting on Quality Levels

Panel A: Quality Metrics

	Control Mean	ITT (1)	TOT (2)
Corners	2.98	1.11 ***	1.70 ***
		(0.12)	(0.11)
Waviness	2.99	1.10 ***	1.68 ***
		(0.12)	(0.10)
Weight	3.08	1.07 ***	1.63 ***
		(0.11)	(0.11)
Touch	3.12	0.40 ***	0.66 ***
		(0.06)	(0.07)
Packedness	3.11	0.89 ***	1.59 ***
		(0.11)	(0.12)
Warp Thread Tightness	3.05	0.83 ***	1.49 ***
		(0.10)	(0.12)
Firmness	2.98	0.87 ***	1.60 ***
		(0.11)	(0.12)
Design Accuracy	3.17	0.79 ***	1.41 ***
		(0.10)	(0.12)
Warp Thread Packedness	3.05	1.07 ***	1.65 ***
		(0.11)	(0.11)
Inputs	3.07	0.89 ***	1.62 ***
		(0.10)	(0.12)
Loom	2.02	0.03	0.05
		(0.02)	(0.04)
R-squared		0.44	0.60
Observations		6,885	6,885

Panel B: Stacked Quality Metrics

	Control Mean	ITT (1)	TOT (2)
Stacked Quality Metrics	2.96	0.79 ***	1.35 ***
		(0.09)	(0.08)
R-squared		0.39	0.54
Observations		6,885	6,885

Notes: Panel A stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality metric indicator, 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 8: Impact of Exporting on Unadjusted Productivity

	Log(Output Per Hour)		Log (Unadjusted TFP)	
	ITT	TOT	ITT	TOT
	(1)	(2)	(1)	(2)
Treatment	-0.24 ***	-0.42 ***	-0.28 ***	-0.50 ***
	(0.09)	(0.16)	(0.09)	(0.16)
R-squared	0.18	0.16	0.26	0.24
Control Group Mean	0.26	0.26	0.49	0.49
Observations	687	687	674	674

Notes: Table reports treatment effects on the two productivity measures: (log) output per hour and (log) unadjusted TFP. See Appendix A for the methodology used to obtain unadjusted 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 9: Step 1: Quality and Productivity

Panel A: Specification Controls

	Stacked Quality Metrics		Log(Output per Hour)		Log(TFP)	
	ITT	TOT	ITT	TOT	ITT	TOT
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.32 ***	0.78 ***	0.18 **	0.44 **	0.14 **	0.35 **
	(0.04)	(0.08)	(0.08)	(0.18)	(0.07)	(0.16)
(log) Thread quantity	0.04	0.02	-0.12	-0.13	-0.07	-0.08
	(0.05)	(0.04)	(0.13)	(0.13)	(0.13)	(0.12)
Difficulty Control	0.47 ***	0.34 ***	-0.14 ***	-0.21 ***	-0.16 ***	-0.22 ***
	(0.02)	(0.03)	(0.04)	(0.05)	(0.04)	(0.05)
(log) # colors	0.03 **	0.01	-0.05 *	-0.07 **	-0.06 **	-0.07 ***
	(0.01)	(0.01)	(0.03)	(0.03)	(0.03)	(0.02)
Low-market Segment	-0.19 ***	-0.08 **	0.43 ***	0.49 ***	0.42 ***	0.47 ***
	(0.03)	(0.03)	(0.08)	(0.09)	(0.07)	(0.08)
Mid-Market Segment	-0.19 ***	-0.05	0.29 ***	0.36 ***	0.26 ***	0.32 ***
	(0.04)	(0.04)	(0.08)	(0.09)	(0.07)	(0.08)
Rug Type FEs	yes	yes	yes	yes	yes	yes
Input Tread Type FEs	yes	yes	yes	yes	yes	yes
R-squared	0.64	0.67	0.57	0.58	0.62	0.63
Observations	6,820	6,820	673	673	660	660

Panel B: Specification Fixed Effects

	Stacked Quality Metrics		Log(Output per Hour)		Log(TFP)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.13 **	0.53 ***	0.31 ***	1.26 ***	0.25 ***	0.94 ***
	(0.05)	0.17	(0.08)	(0.45)	(0.08)	(0.36)
Specification FEs	yes	yes	yes	yes	yes	yes
R-squared	0.71	0.72	0.69	0.62	0.70	0.68
Observations	6,820	6,820	428	428	416	416

Panel C: Specification-Adjusted Dependent Variables

	Stacked Quality Metrics		Log(Output per Hour)		Log(TFP)	
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment	0.42 ***	0.72 ***	0.18 **	0.33 ***	0.25 ***	0.45 ***
	(0.05)	(0.04)	(0.07)	(0.13)	(0.07)	(0.12)
R-squared	0.18	0.27	0.06	0.10	0.08	0.14
Observations	6,860	6,860	678	678	671	671

Notes: Table reports treatment effects on the stacked quality measures, and the two productivity measures. The TOT specifications instrument takeup with treatment. There are 7 rug type fixed effects. In addition to the controls shown in the table, the regressions also control for baseline values of the dependent variable, as well as round, strata and rug type fixed effects. For the TFP regressions, Panels A and B use the unadjusted TFP measure and Panel C uses the specification-adjusted TFP measure; see Appendix A for details. Panel C also uses specification-adjusted variants of the quality metrics and log output per hour. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 10: Step 2: Quality and Productivity on Identical Domestic Rugs

Panel A: Quality Metrics

	Master Artisan			Professor		
		ITT	TOT	Control	ITT	TOT
	Control Mean	(1A)	(1B)	Mean	(2A)	(2B)
Corners	3.23	0.72 ***	1.05 ***	3.31	0.29 **	0.45 **
		(0.14)	(0.17)		(0.13)	(0.18)
Waviness	3.17	0.55 ***	0.83 ***	3.31	0.25 **	0.36 **
		(0.14)	(0.18)		(0.12)	(0.17)
Weight	3.60	0.62 ***	0.91 ***	3.64	0.58 ***	1.01 ***
		(0.13)	(0.17)		(0.17)	(0.27)
Packedness	3.30	0.77 ***	1.10 ***	3.28	0.28 **	0.43 ***
		(0.13)	(0.16)		(0.11)	(0.16)
Touch	3.29	0.52 ***	0.79 ***	3.27	0.36 ***	0.53 ***
		(0.11)	(0.14)		(0.12)	(0.17)
Warp Thread Tightness	3.00	0.51 ***	0.74 ***	3.30	0.25 **	0.39 **
		(0.09)	(0.12)		(0.12)	(0.17)
Firmness	3.21	0.71 ***	1.01 ***	3.23	0.29 **	0.43 **
		(0.14)	(0.18)		(0.12)	(0.17)
Design Accuracy	3.65	0.53 ***	0.83 ***	3.45	0.27 **	0.39 **
		(0.11)	(0.16)		(0.11)	(0.16)
Warp Thread Packedness	3.05	0.87 ***	1.28 ***	3.20	0.39 ***	0.62 ***
		(0.14)	(0.18)		(0.12)	(0.17)
R-squared		0.21	0.32		0.11	0.11
Observations		1,680	1,680		1,667	1,667

Panel B: Stacked Quality Metrics

	Master Artisan			Professor		
		ITT	TOT	Control	ITT	TOT
	Control Mean	(1A)	(1B)	Mean	(2A)	(2B)
Stacked Quality Metric	3.28	0.64 ***	0.94 ***	3.33	0.33 ***	0.48 ***
		(0.10)	(0.12)		(0.10)	(0.13)
R-squared		0.19	0.32		0.09	0.13
Observations		1,680	1,680		1,667	1,667

Panel C: Additional Quality Metrics

		ITT	TOT
	Control Mean	(1A)	(1B)
Length Accuracy	-4.51	1.43 ***	1.97 ***
		(0.51)	(0.75)
Width Accuracy	-2.29	0.17	0.26
		(0.29)	(0.45)
Weight Accuracy	-221.0	89.1 ***	148.0 ***
		(20.3)	(32.2)
Time (in minutes)	247.0	-5.67	-8.9
		(6.6)	(9.7)
R-squared		0.84	0.82
Observations		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 objective accuracy measures which are calculated as the negative of the abosulte error for the specification, so that 0 is perfect and a higher value is better. It also includes 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 11: Step 4: Information Flows and Quality Levels

	Stacked Quality Metrics (1)	Specification Adjusted Stacked Quality Metrics (2)
$Takeup_i$ x {Talked About Dimension}$_{id}$	0.15 ***	0.13 **
	(0.05)	(0.05)
Quality Metric FEs	yes	yes
$Takeup_i$ x Quality Metric FEs	yes	yes
Firm-fixed Effects	yes	yes
Specification-adjusted Quality Metrics	no	yes
R-squared	0.81	0.56
Observations	1,700	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. Column 2 uses the specification-adjusted quality metrics described in the text. Note that the regressions with specification controls and specification fixed effects are identical to the results in Column 1 because the firm fixed effects absorb all the variation in specifications. Regressions are run on a cross-section of firms and include baseline values, firm fixed effects, quality metric fixed effects, and quality metric fixed effects interacted with takeup. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Figure 1: Example of Rug Specifications Provided by Potential Foreign Client

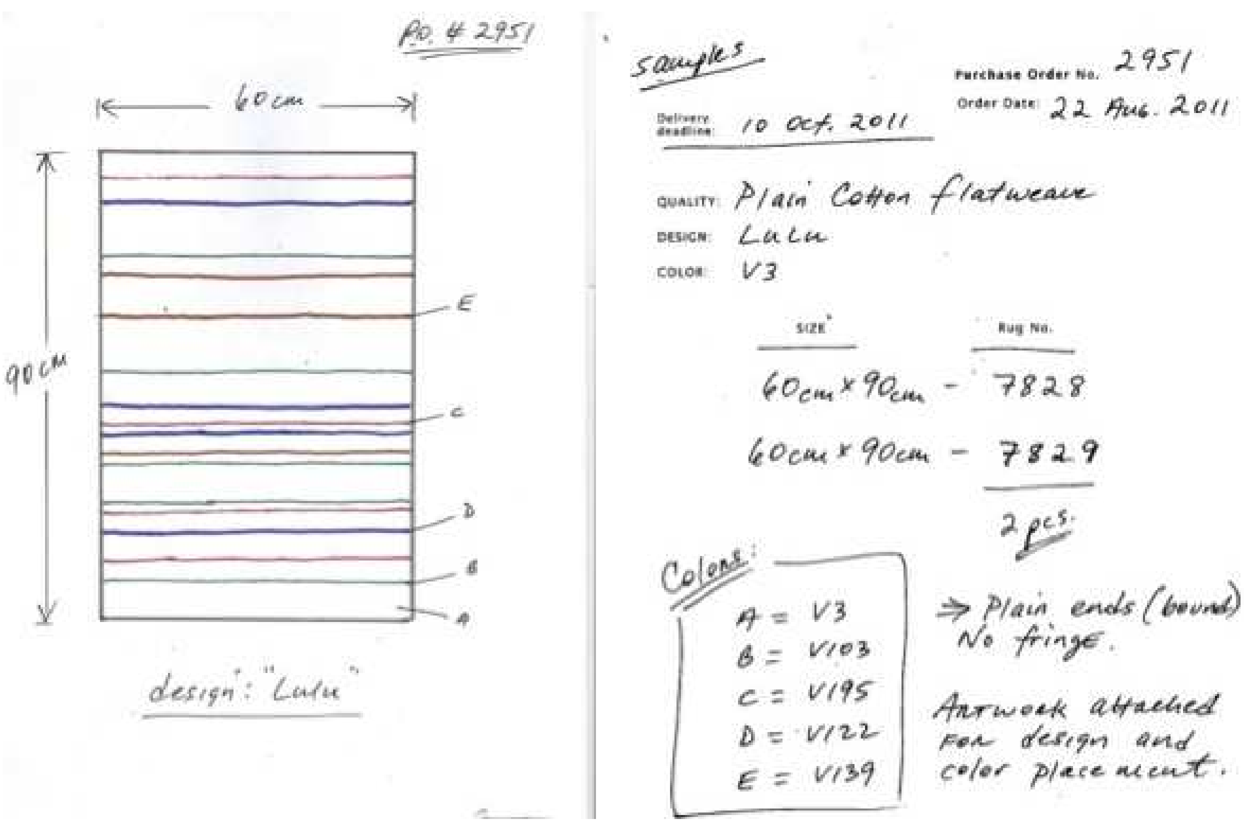


Figure 2: Example of Domestic Rugs and Export Rugs

Notes: Figure provides examples of a domestic rug (left) and export rugs ordered by foreign buyers (center, right). The domestic rug is the duble rug that firms were asked to manufacture at the Quality Lab (See Section 6.3).

Figure 3: Cumulative Export Orders

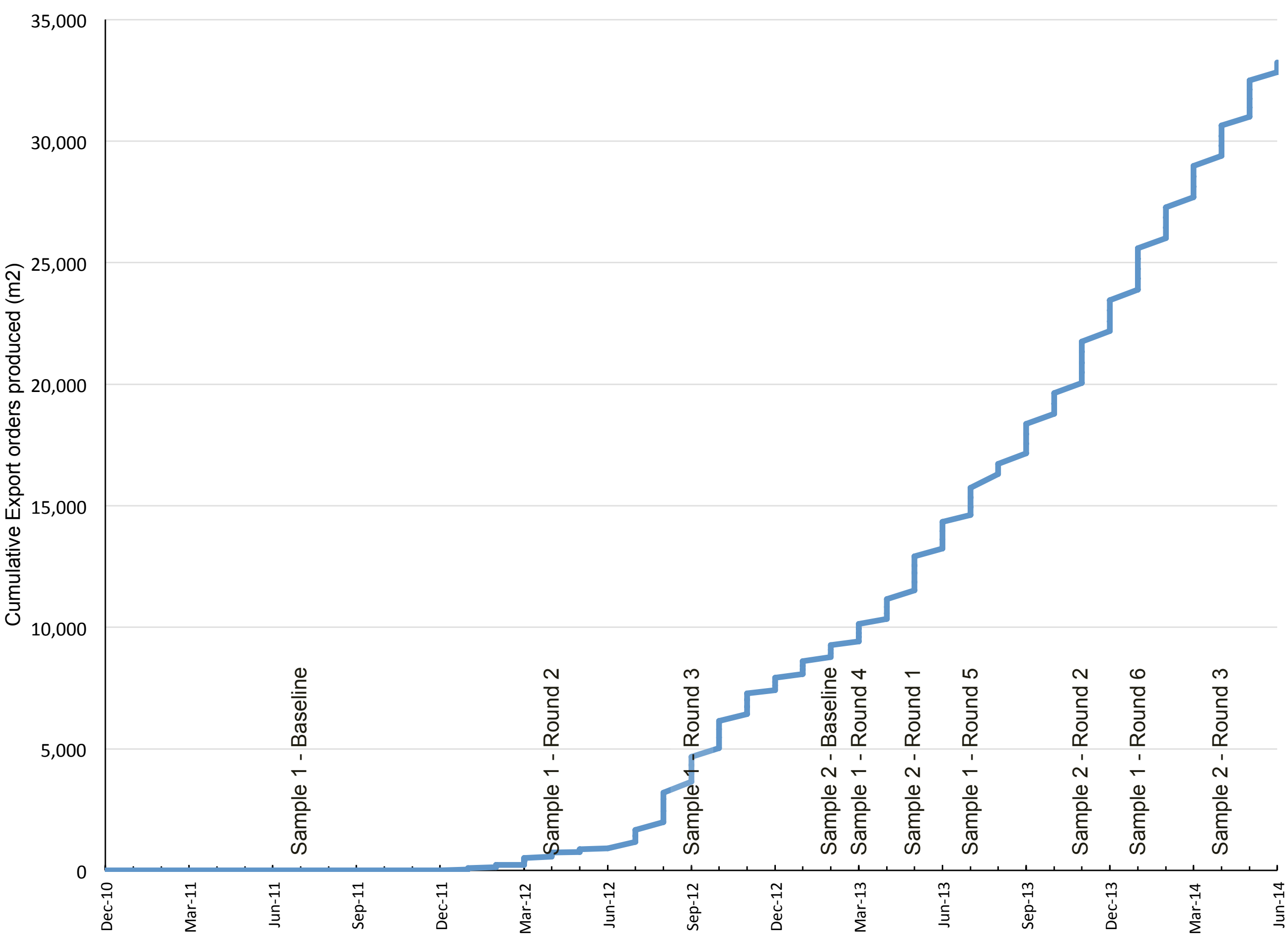
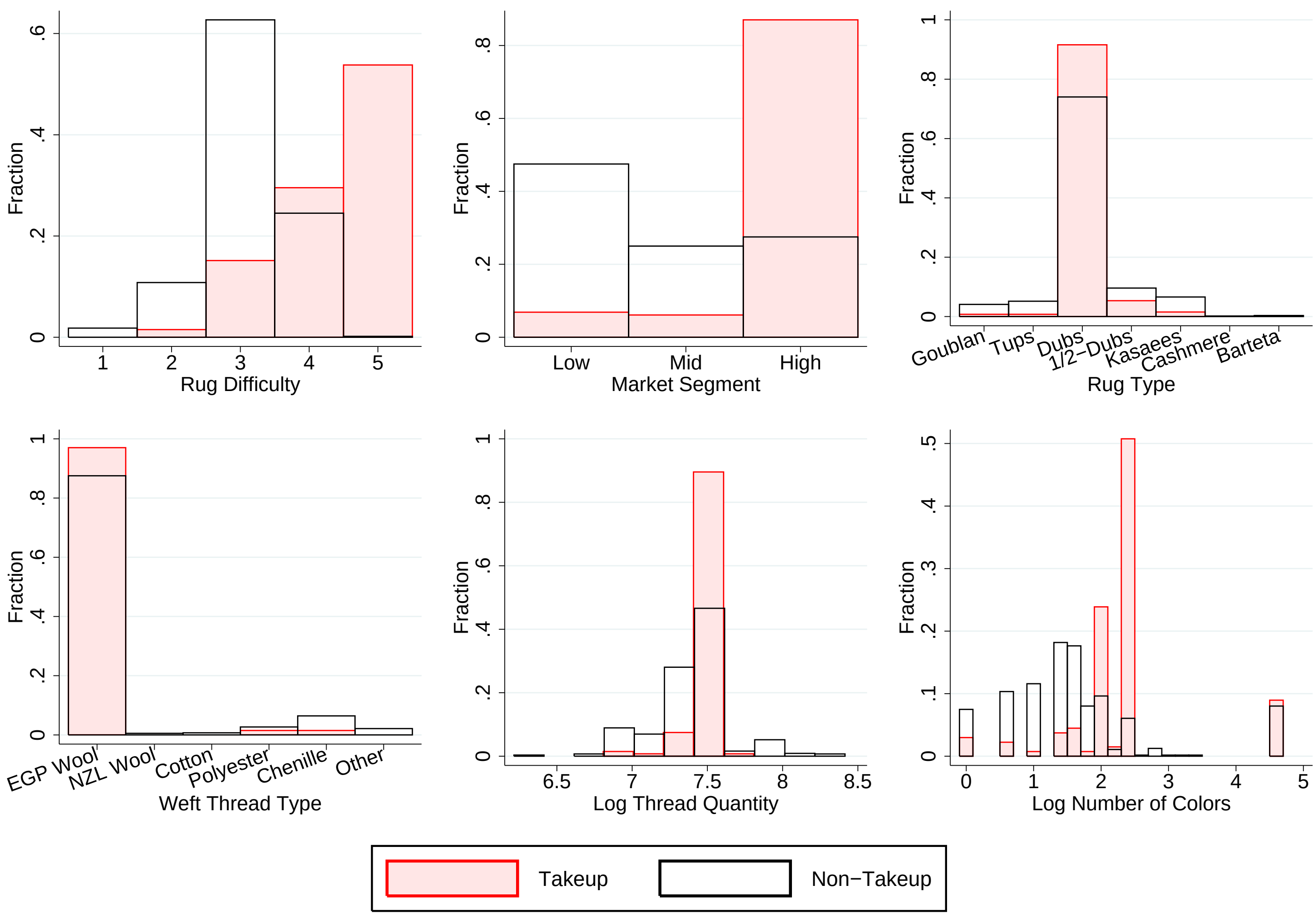
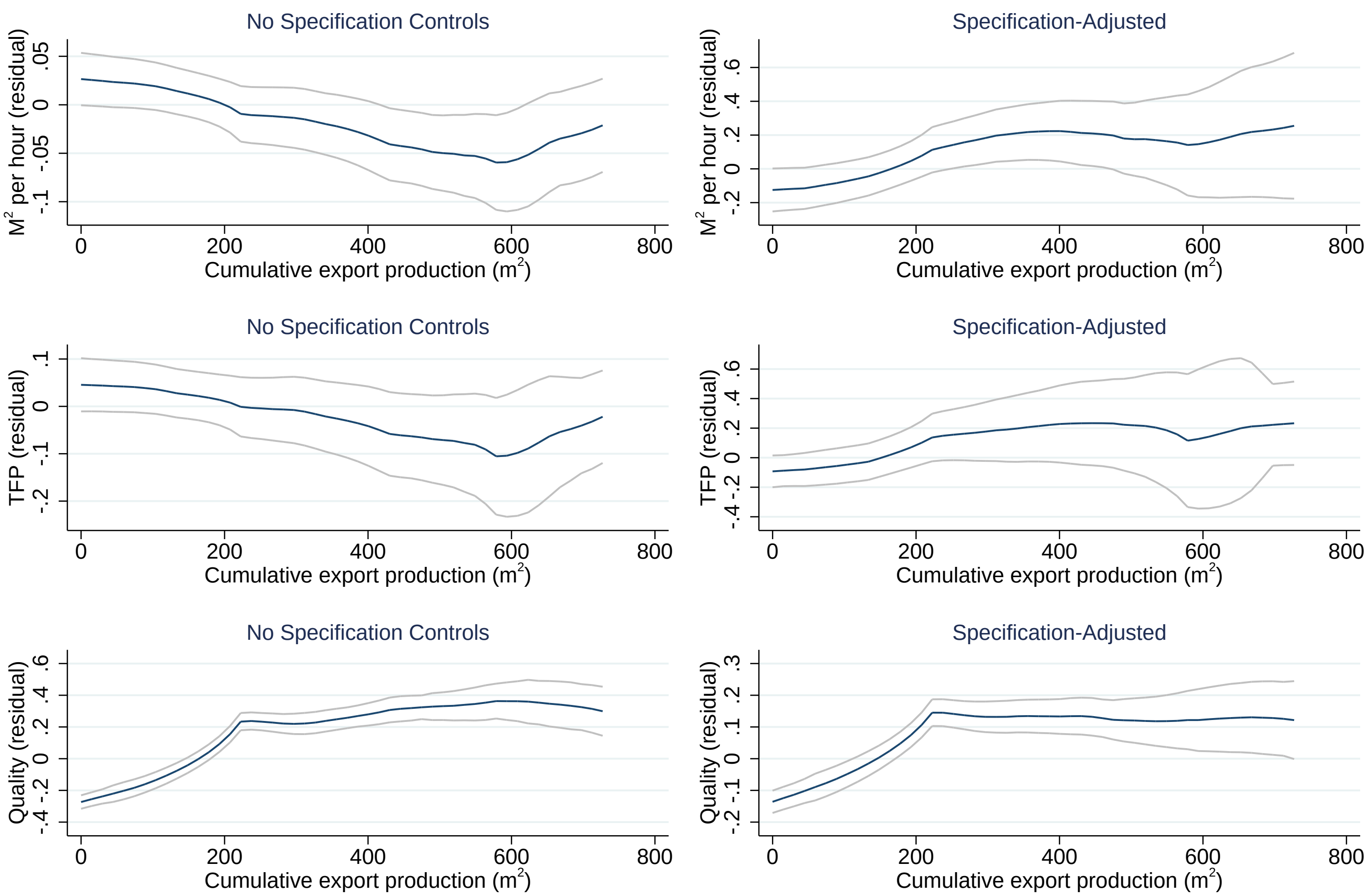


Figure 4: Overlap in Rug Specifications on Domestic and Export Orders



Notes: Figure plots the density of the six rug specifications for takeup (shaded) and non−takeup (outline) firms.

Figure 5: Step 3: Learning Curves, Takeup Firms



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

Exporting and Firm Performance: Evidence from a Randomized Trial

Appendices for Online Publication

Online Appendix B The Global and Egyptian Rug Industry

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).[3] 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, and in 2013 Egypt's exports in HS57 ("carpets and other textile floor coverings") constituted 1.4 percent of its exports. The top 5 destinations for these exports are the U.S., Germany, Italy, United Kingdom, and Canada, which account for 59 percent of total exports in the category.

Online Appendix C The Rug Production Process in Fowa

The firms in Fowa typically operate out of a rented space or sometimes the home of the owner. The owner is almost always the primary weaver and most firms have no other full time employees. Firms self-identify as specialists in one of four flat-weave rug types: duble, tups, kasaees and goublan. 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 Figure C.1 for pictures.

The process of producing rugs is standardized across firms. The elements of the production technology are marked in Figure C.2. 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.

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

Duble rugs are the thickest of the rug types and typically made using a reed with 200 openings per meter. In contrast, tups and goublan rugs are more intricate. Goublan rugs usually require a reed with 400 openings per meter and because of the intricate design patterns, weavers use their hands, instead of shuttles, for precision in weaving the weft thread. Tups rugs are also finely woven using a reed with 400 openings per meter but because they are used as floor rugs and have simpler designs, weavers use shuttles to increase the pace of production. Finally, kasaees rugs use reeds with 250 openings per meter but these rugs use left-over cloth (for example, torn up t-shirts) in their weave instead of more expensive wool inputs.

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

Notes: Figure illustrates the four flat-weave rug types produced by firms in Fowa, Egypt: duble, tups, kasaees, and goublan.

Figure C.2: Production Technology

Online Appendix D Additional Details on Experiment Protocol

In this appendix, we provide a detailed description of the experiment protocol. For completeness, we repeat the details reported in the main text, as well as provide additional information.

Online Appendix D.1 Experiment Protocol

Hamis Carpets (with ATA's assistance) marketed rugs to overseas buyers. Once export orders were secured, we divided the orders into smaller packets. The treatment firms were visited by our survey team and a representative of Hamis carpets and offered the opportunity to produce one of these packets of export orders. More precisely, Hamis Carpets showed them the rug design, explained 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 by trading off the desire to have a reasonable sized initial order and the need to have enough orders to treat the firms. Hamis was given discretion regarding what price to offer the treatment firms, and chose prices based on the specifications of the particular rug orders. The initial protocol intended for us to offer this packet in one go, and for the rug orders to be of the type the firm was specialized in producing. As we discuss in Online Appendix D.2, this was the protocol followed for Sample 2. Given our initial difficulties in generating export orders, in Sample 1 we could only offer Duble and Kassaes rugs orders, and only by sequentially offering smaller packet sizes that summed to 110m^2.[4]

If the firm accepted, Hamis delivered the input thread and the correctly sized reed and heddle

[4] As Kassaes rugs are substantially quicker to produce we offered an initial packet size of 250m^2 to Kassaes producers.

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 and receive 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. The experiment protocol simply forced Hamis to offer an *initial* order to the treatment firm.[5] 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.[6] Thus, the intervention provided treatment firms with the opportunity to produce rugs for the export market.

Online Appendix D.2 Sample, Randomization and Takeup

Since there is no official census of firms that manufacture rugs in Fowa, we hired a Fowa-based NGO to go street by street looking for rug-making firms in homes and workshops that satisfied the following characteristics: a) had fewer than 5 employees; b) worked on their own account (meaning that they bought their own inputs); c) had never previously worked with Hamis, d) were willing to participate in a research experiment on exporting. Based on our power calculations, we stopped the recruitment drive when the NGO had identified 300 firms. As there are far more than 300 rug-making firms in Fowa, the first recruitment drive provided us with a sample of firms in central areas of Fowa that were operating and observable at the time of recruitment. This exercise produced a list of 303 firms, which we refer to as "Sample 1".

Firms specialize in one of four rug types. We stratified the sample on the type of rug produced and the loom size. Within each stratum we randomized firms into treatment and control using a random number generator and strata that contained an odd number of firms were assigned one more firm to control than to treatment. For reasons that will become clear momentarily, we refer to these 303 firms as "Sample 1". The first two rows of columns 2, 3, 4 and 6 in Table 1 show the total number of firms by rug type and treatment status for Sample 1.

The third row of Table 1 shows the takeup status for these firms. 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, takeup rates were 10 and 19 percent, respectively. We expected low takeup 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. In follow-up round 2, we

[5]In cases in which the treatment firms were unable to produce the rugs at the quality level required by Hamis Carpets, the research team would reimburse Hamis carpets the cost of the materials that were used in the first 110m^2. This happened in only 3 cases, and in none of Duble producing-firms that comprise the joint sample we use in the paper (see Table 1).

[6]As noted in the draft, 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.

asked all treated firms who did not takeup why they did not takeup. Table E.1 presents these results and confirms that the main reason for refusals among goublain and tups firms was that the export rug order was not the suitable rug type.

Table 1 shows that among kasaees and duble rug producers, takeup was 26 and 38 percent, respectively. These takeup rates were still relatively low. As we detail in Section 2.2, between December 2010 and May 2012 we were able to generate a small number of export orders for kasaees and duble rugs (see Figure 3). However, given the small number of export orders, we were unable to approach treatment firms in Sample 1 with the opportunity to produce the full 110m^2 in one go. Instead, we had to sequentially offer smaller orders of 10-20m^2, or about two weeks of work. Because these orders were small, many firms were unwilling to work with us: Table E.1 shows that the most common reason why duble firms did not take up was an unwillingness to jeopardize their existing relationships with other intermediaries for a small amount of work, and that the export order was not the suitable rug type (primarily firms that 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.

From March 2013, Hamis' major buyers offered assurances that they would continue to place duble rug orders for the foreseeable future so it became feasible to offer the opportunity to produce 110m^2 in one go. Since this was the intended treatment in our experimental protocol, we conducted a second round for recruitment of firms that specialized in duble rugs. At this stage of the project, we had an experienced Fowa-based team who was able to locate an additional 140 firms that had been not been captured at the initial recruitment drive in July 2011. We refer to these additional firms as "Sample 2" and again stratified these firms by loom size. We randomized 35 firms in the treatment group using a random number generator. The choice of 35 treatment firms for Sample 2 was dictated both 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. Column 5 of Table 1 reports treatment and takeup statistics for Sample 2. Consistent with the low takeup in Sample 1 being due to the inability to offer the full 110m^2 packet in one go, 32 out of 35 treatment firms in Sample 2 took up the opportunity to produce export orders.[7]

As noted, the experiment protocol allowed Hamis Carpets to exit their relationship with firms that did not produce the 110m^2 at the expected quality level and time frame. The 5th row of Table 1 reports the number of "successful" takeup firms, defined as those who produced more than 110m^2 and received subsequent orders from Hamis. 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.

Samples 1 and 2 do differ on some observables. On average, Sample 2 firms have higher profits and consumption, although similar quality levels and output. These differences may be due to the fact that recruitment was carried out sequentially, so that Sample 2 firms were more diffi-

[7] 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.

cult to find or less centrally located. However, there are no significant differences in observables between treatment and control groups within either sample as can be seen in Table F.1, or in the joint sample of Duble firms as can be seen in Table 3. We also find that the estimated Treatment of the Treated (TOT) effects do not differ across samples (we compare TOTs as unlike ITTs they account for the very different takeup rates across the two samples): The TOT estimates are only statistically different from each other at the 10 percent level for 3 of the 34 results reported in the paper (none are significant at the 5 percent level).

Online Appendix D.3 Protocol for Quality Lab

In June 2014, we rented a workshop with a loom and invited the primary weaver in all our treatment and control firms to come to the workshop to produce an identical rug. The rug specifications mimicked a popular rug design sold at mid-tier domestic retail outlets in Egypt (a 140cm by 70cm rug with a desired weight of 1750g, and the master artisan assigned a difficulty rating of 3—see Figure 2 for the actual design). Prior to the firm's weaver (who was most often the owner of the firm) arriving, our field team prepared the loom for weaving with the help of the master artisan. We provided all the necessary thread, reed and heddle inputs. Thus, each firm faced an identical loom, an identical loom setup, an identical set of inputs, and were asked to manufacture identical rugs.

This setting is not unusual in Fowa since a handful of large intermediaries have workshops where they invite producers to weave. Nonetheless, our Quality Lab differed in that we invited firms to produce a single rug, instead of many meters of rugs as is the market norm. To mitigate potential Hawthorne effects, where the firms act differently because they know they are being analyzed, we excluded Hamis Carpets from being involved in administering the Quality lab, from participating in contacting the firms, and from encouraging them to participate. Instead, we hired a new staff member to administer the Quality Lab, and he was instructed to invite sample firms to produce a rug at the rented workshop for a LE70 payment, which is about three times the average price for a rug of these specifications. The firms were told that the reason why this lucrative opportunity existed was because there was a buyer in Cairo that was interested in purchasing one rug from many different producers in Fowa. This claim was truthful, as one of the authors was in Cairo at the time, serving the role of that buyer.

The firms were shown to the Quality Lab and provided the design and specifications of the rug. The firms were not incentivized to produce quality in any way, nor were they encouraged to produce quickly or slowly or given a time limit. They also were not given any reason to believe that their performance on this task would impact future opportunities to produce for this buyer, or given any information about the desired quality of the rug. They were not even told they were being evaluated in any way. Firms were simply informed that as long as they completed the rug they would be given LE70 as promised.

The total time taken to weave the rug was timed and after the rug was completed, the staff member recorded the length, height and weight of the rug, and tagged each rug with a new set of firm identification numbers. After all firms had manufactured the rugs, they were sent to the master artisan to score the quality levels without knowing the identity of the firm that made them. In addition, the rugs were sent to Professor Fayrouz Al-Gamal, a Professor of Handicraft Science from Domietta University, to provide an independent set of quality scores. Professor Al-Gamal has been the chair of the "Spinning, Weaving and Knitting Department" at the Domietta University since 2013 and is an expert in Jacquard knitting techniques. He has contributed technical chapters to eight published books, and presented his work at twelve academic conferences over the last five years.

Online Appendix E Additional Tables and Figures

Table E.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	6.1	6	14.3	5	26.3	15	38.5	28	18.8
Risk relationship with current intermediary	2	4.1	1	2.4	2	10.5	7	17.9	12	8.1
Price was too low	2	4.1	1	2.4	2	10.5	3	7.7	9	6.0
Left industry or passed away	2	4.1	3	7.1	3	15.8	5	12.8	13	8.7
Export order not suitable rug type	39	79.6	30	71.4	6	31.6	7	17.9	82	55.0
Refused contact with survey team	1	2.0	1	2.4	1	5.3	2	5.1	5	3.4
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 E.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 E.3: Geographic Spillovers to Control Firms

	Sum of Inverse Distance to Treatment Firms	Sum of Inverse Squared Distance to Treatment Firms	Marginal Effect	R-Squared	Obs.
Ever Exported	-0.16	16.29	6.78	0.12	128
	(0.92)	(40.24)	(16.40)		
Direct Log Monthly Profits	-0.89	0.28	-0.77	0.26	368
	(1.35)	(43.17)	(17.22)		
Direct Log Profits per Hour	-0.50	-40.50	-17.72	0.18	368
	(1.43)	(47.51)	(18.97)		
Log Output per Hour	-1.28	37.33	19.11	0.08	427
	(0.99)	(36.91)	(16.36)		
Stacked Quality	0.60	-7.09	-2.41	0.06	4,408
	(0.41)	(15.91)	(6.43)		
Log Unadjusted TFP	-1.53	103.00 *	42.00 *	0.22	418
	(1.86)	(59.27)	(23.85)		
Log Adjusted TFP	-1.71 *	55.38	21.85	0.10	-421
	(0.96)	(42.50)	(17.30)		

Notes: Table reports results from regressing the outcome variables in each row on inverse distances between control firms and all treatment firms and an inverse distance squared term (measured in meters). The third column shows the marginal effect of distance on the outcome based on the results from the regression. Regressions include round and strata fixed effects and control for baseline values. The stacked quality regression includes metric fixed effects. Standard errors clustered at the firm level. Significance * .10; ** .05; *** .01.

Table E.4: Key Results using Firm Fixed-Effects

	ITT (1)		TOT (2)		R-squared (3)	Obs. (4)
Ever Exported	0.53	***	0.74	***	0.71	410
	(.12)		(.10)			
Direct Log Monthly Profits	0.30	**	0.47	***	0.57	874
	(.12)		(.16)			
Direct Log Profits per Hour	0.28	**	0.44	***	0.48	874
	(.12)		(.16)			
Log Output per Hour	-0.18		-0.31		0.04	901
	(.12)		(.20)			
Log Stacked Quality	1.04	***	1.58	***	0.60	8931
	(.10)		(.09)			
Log Unadjusted TFP	-0.26	**	-0.44	**	0.04	890
	(.12)		(.20)			
Log Adjusted TFP	0.01				0.04	301
	(.20)					

Notes: Table reports treatment effects on 8 main outcomes for the duble firms in the joint sample using firm fixed effects instead of controling for the baseline value of the variable. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table E.5: Step 3: Treatment Dynamics

	Log(Direct Profits)	Log(Direct Profits per Hour)	Log(Output per Hour)
	(1)	(2)	(3)
Treatment	0.28 ***	0.23 ***	-0.26 **
	(0.08)	(0.07)	(0.11)
x 2nd Round Since Treatment	-0.03	-0.06	0.16
	(0.10)	(0.09)	(0.12)
x 3rd Round Since Treatment	-0.07	-0.12	-0.05
	(0.10)	(0.09)	(0.13)
x 4th Round Since Treatment	-0.07	-0.04	-0.26
	(0.16)	(0.15)	(0.24)
x 5th Round Since Treatment	0.27 *	0.27 *	0.17
	(0.16)	(0.15)	(0.24)
R-squared	0.22	0.15	0.18
Observations	573	573	687

	Log (Unadjusted TFP)	Log (Specification Adjusted TFP)	Log (Specification Adjusted Quality)
	(4)	(5)	(6)
Treatment	-0.29 ***	0.25 **	0.44 ***
	(0.10)	(0.10)	(0.06)
x 2nd Round Since Treatment	0.18	0.08	0.01
	(0.12)	(0.11)	(0.06)
x 3rd Round Since Treatment	-0.08	-0.01	0.01
	(0.12)	(0.11)	(0.07)
x 4th Round Since Treatment	-0.31	-0.29	-0.10
	(0.23)	(0.19)	(0.11)
x 5th Round Since Treatment	0.13	0.00	-0.21 *
	(0.21)	(0.15)	(0.11)
R-squared	0.27	0.09	0.18
Observations	674	671	6860

Notes: Table reports treatment effects interacted with dummies for each round of data collection. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Control group means are reported in levels. Standard errors are clustered by firm. Significance * .10; ** .05; ***

Table E.6: Step 4: Summary of Information Flows

	(1)	
Number of Visits	11.0	
	(2.57)	
Length of Visit (in minutes)	27.6	
	(4.88)	
Discussed technique?	90.3%	
	Discussed Metric?	Discusssed Technique?
	(1A)	(1B)
Corners	31.8%	100.0%
Waviness	20.5%	100.0%
Weight	54.5%	92.9%
Touch	11.4%	100.0%
Packedness	20.5%	93.8%
Warp Thread Tightness	47.7%	78.9%
Firmness	31.8%	100.0%
Design Accuracy	50.0%	96.2%
Warp Thread Packedness	22.7%	75.0%
Observations	44	

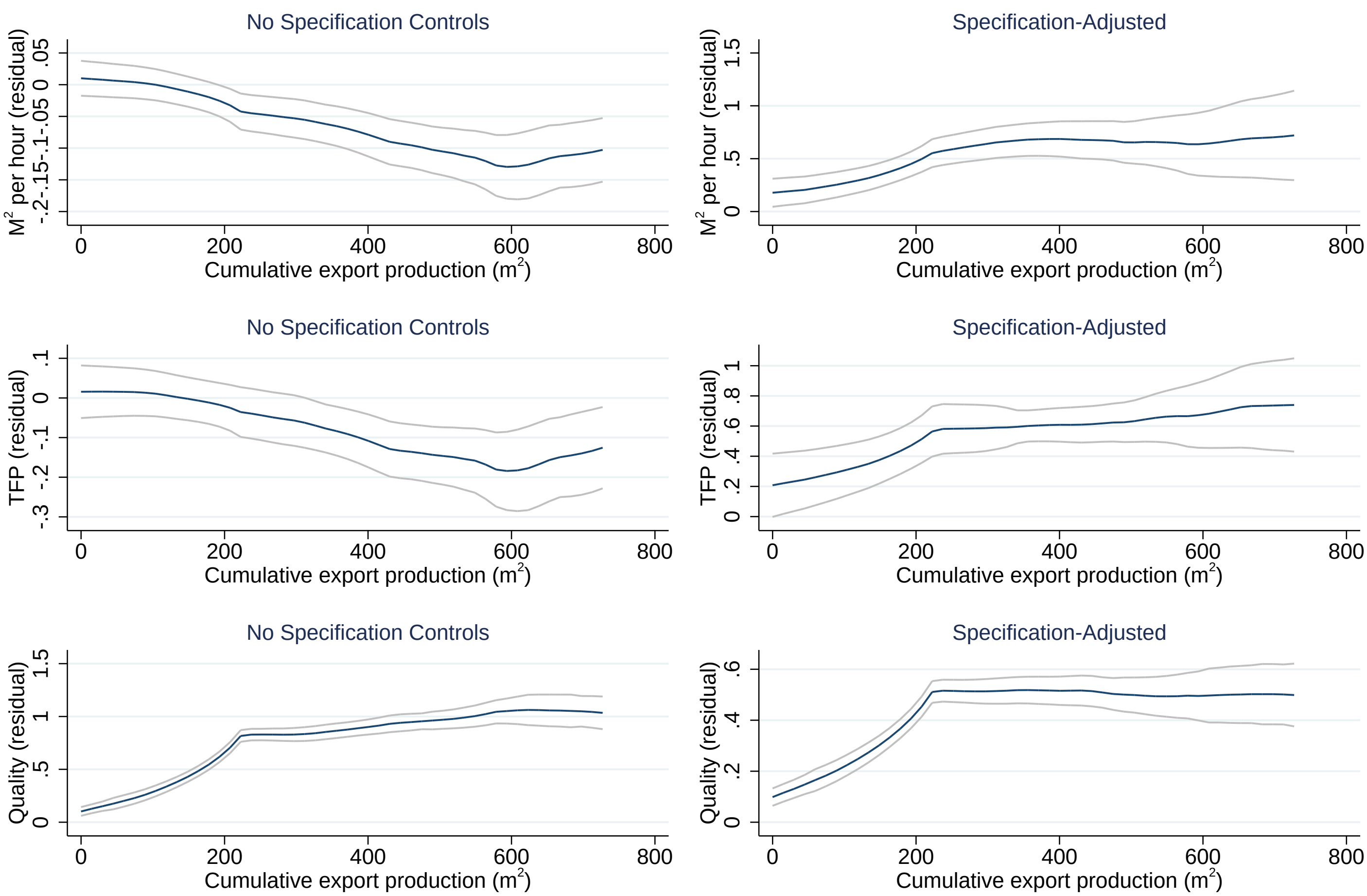
Notes: Table summarize the visits between the intermediary with firms. All firms were visited at least 7 times, and the top panel reports the average length of visit in minutes (with standard deviations in parantheses), and the proportion of interactions overall that discuss technique, rather than simply 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 E.7: Step 5: Quality Hedonic Regression

	Log Profits Per Hour	
	(1)	
Corners	0.08	
	(0.06)	
Waviness	-0.044	
	(0.06)	
Weight	-0.027	
	(0.05)	
Touch	0.189	**
	(0.07)	
Packedness	-0.150	**
	(0.07)	
Warp Thread Tightness	0.173	**
	(0.08)	
Firmness	-0.103	
	(0.08)	
Design Accuracy	0.106	*
	(0.05)	
Warp Thread Packedness	0.032	
	(0.06)	
P-Value of Joint F-Test	0.013	
R-squared	0.587	
Observations	563	

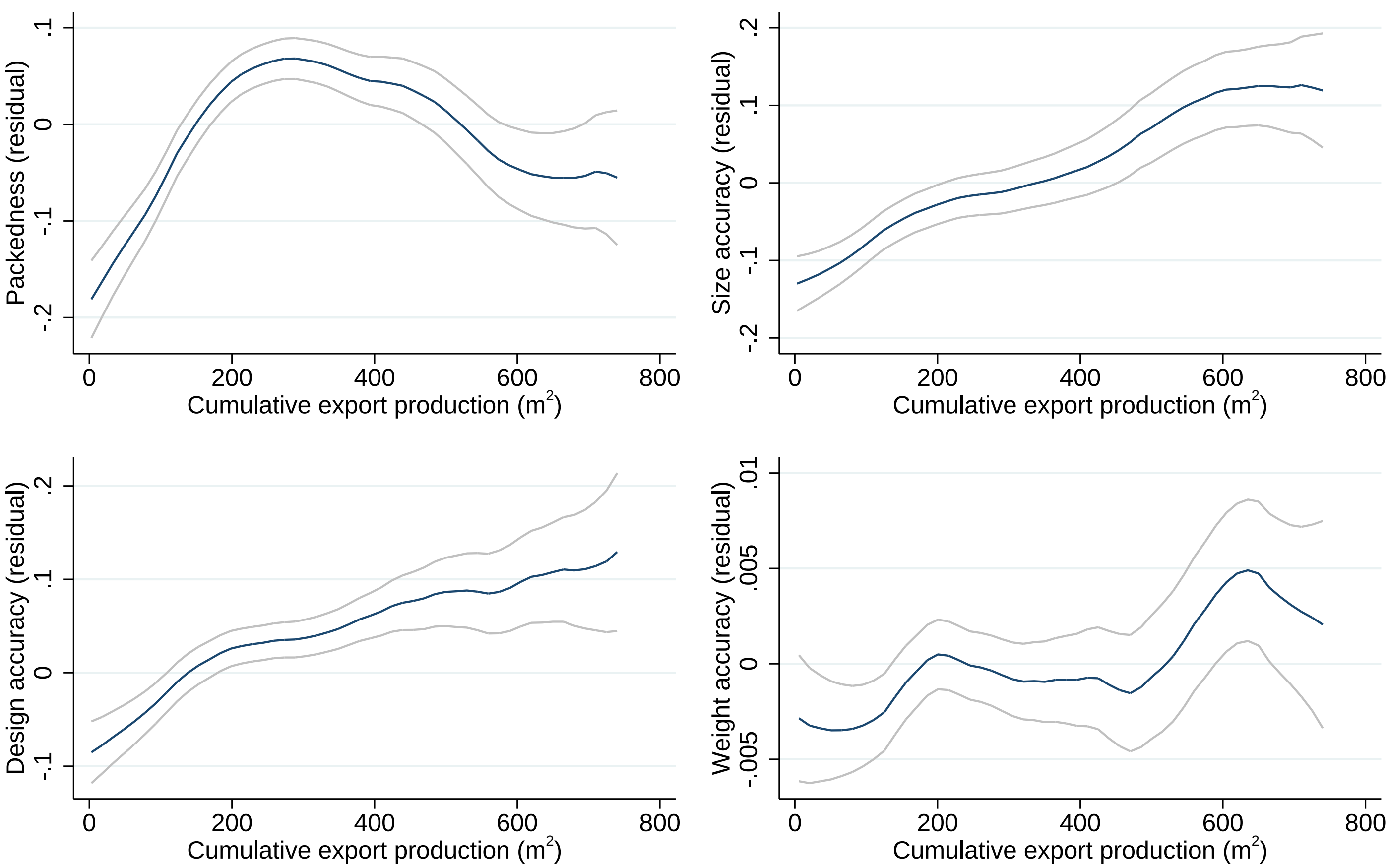
Notes: Table reports the regression of log profits per hour on nine quality metrics (loom and input metrics are excluded since they were held fixed across firms and hence not measured in Step 2, as noted in the text). The regression is run on firms producing for the domestic market (i.e non-takeup firms). The regressions include the same controls for rug specifications used in Step 1, as well as strata and round fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Figure E.1: Step 3: Semi-Parametric Learning Curves, Takeup Firms



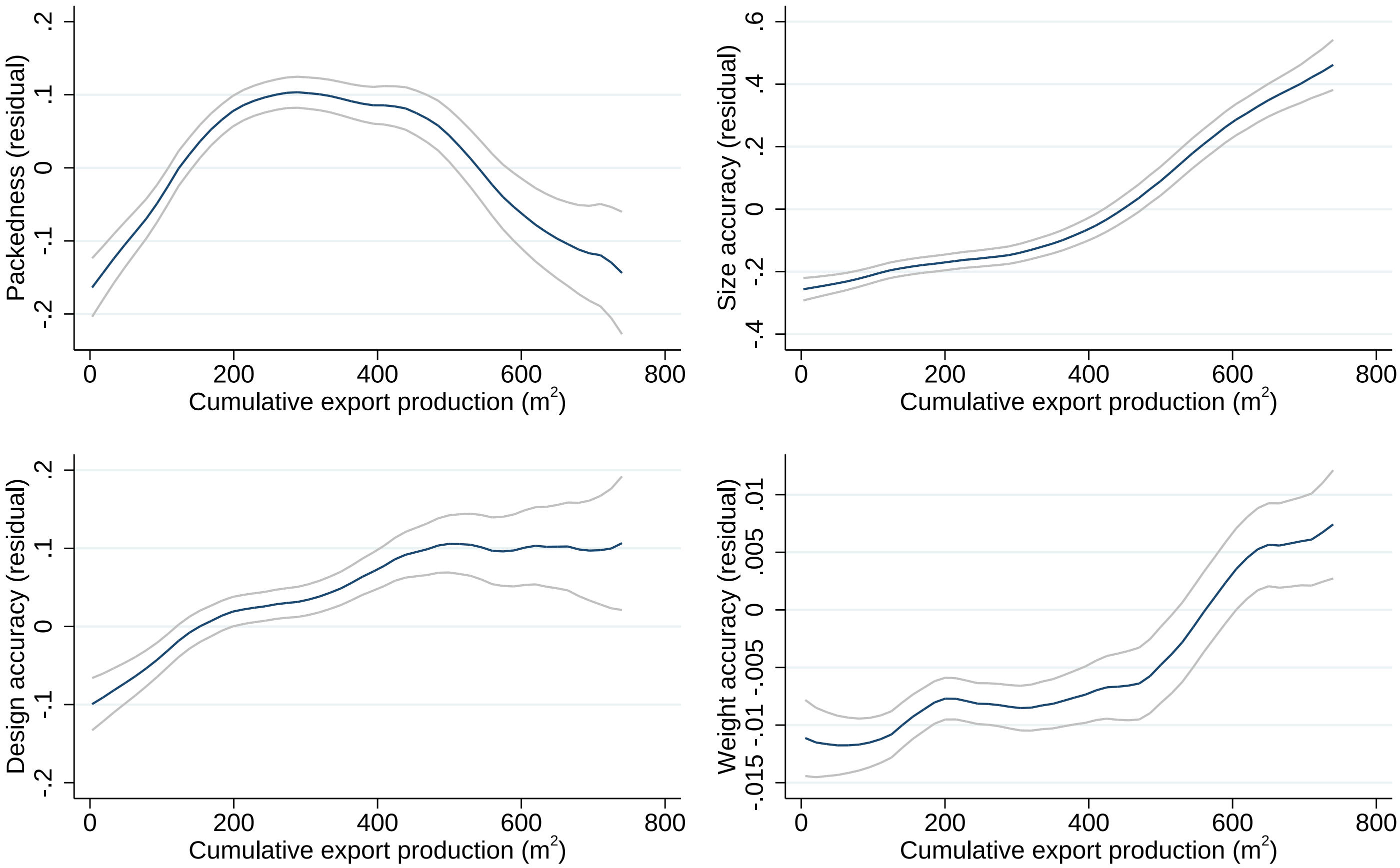
Notes: Figure plots learning curves for quality and productivity measures removing firm and round fixed effects using the partially linear panel data estimator proposed by Baltagi and Li (2002). We present non-parametric plots of the residuals against cumulative export production for takeup firms. Figure restricts attention to range of cumulative exports achieved by Sample 2 firms.

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



Notes: Figure plots learning curves obtained by regressing high-frequency quality measures on firm fixed effects and then plotting a kernel-weighted local polynomial of the residuals against cumulative export production for takeup firms. Quality measures are recorded by the intermediary firm Hamis Carpets for each batch of rugs delivered by each firm (often at a weekly frequency). 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. Figure restricts attention to range of cumulative exports achieved by Sample 2 firms.

Figure E.3: Step 3: Semi-Parametric Learning Curves using High-Frequency Order-Book Data, Takeup Firms



Notes: Figure plots learning curves for high-frequency quality measures removing firm fixed effects using the partially linear panel data estimator proposed by Baltagi and Li (2002). We present non-parametric plots of the residuals against cumulative export production for takeup firms. Quality measures are recorded by the intermediary firm Hamis Carpets for each batch of rugs delivered by each firm (often at a weekly frequency). 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. Figure restricts attention to range of cumulative exports achieved by Sample 2 firms.

Figure E.4: Step 3: Total Export Production CDFs

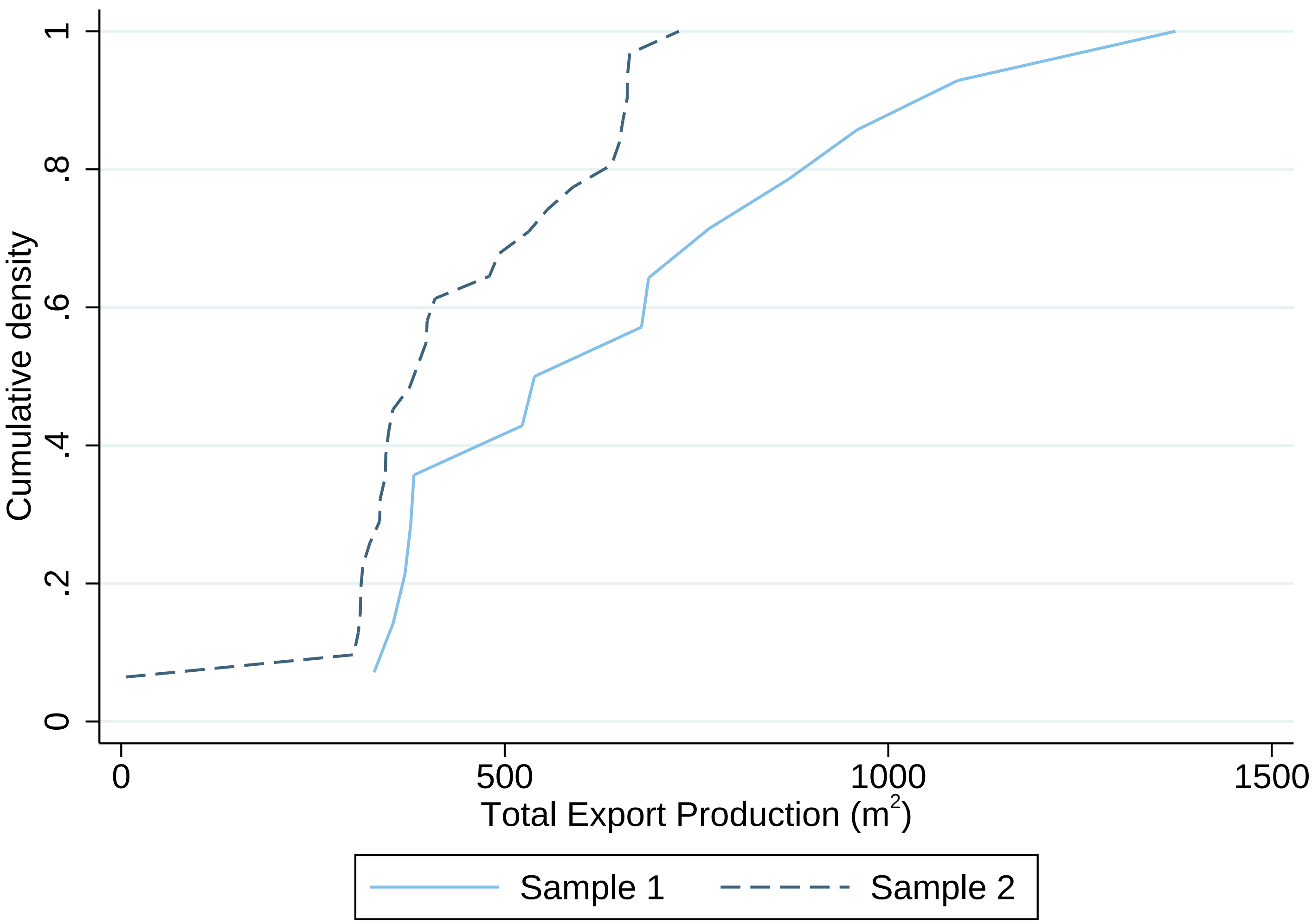


Figure E.5: Step 3: Cumulative Exports and Days Since First Order

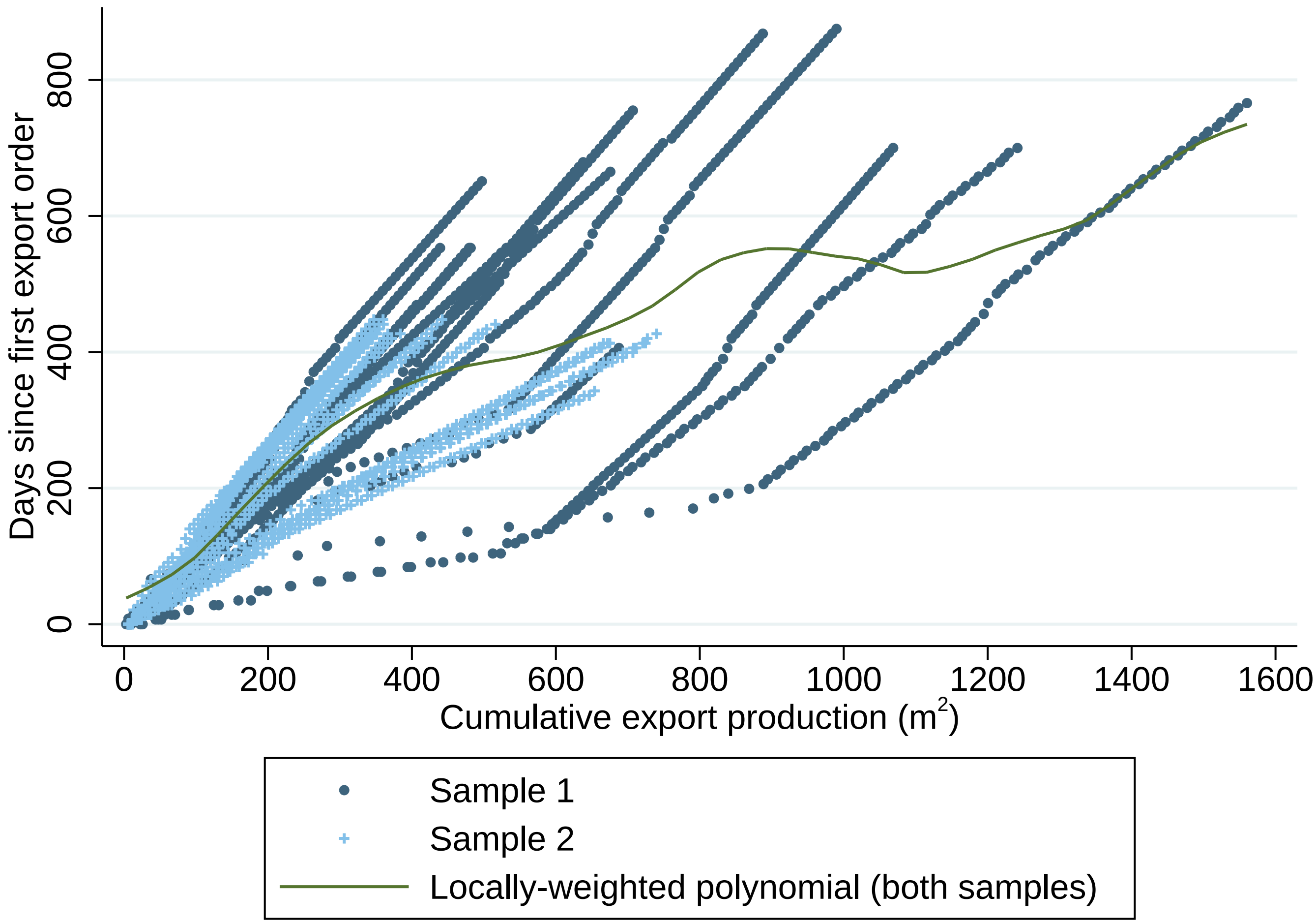


Figure E.6: Step 4: Quality Problems Noted by Overseas Buyer

Online Appendix F Results for Sample 1 duble firms, Sample 2 firms, and Sample 1 non-duble firms

Tables F.1 to F.10 and Figures F.1 to F.2 in this appendix repeat the key tables and figures from the main text but now splitting the Joint Sample into Sample 1 duble firms and Sample 2 firms (recall all Sample 2 firms are duble firms). Table F.11 reports the treatment effects for the non-duble firms in Sample 1 who we were not able to provide with sufficient orders in their specialist rug type. We only have 3 additional rounds of data for these non-duble firms as we stopped surveying them when it became apparent we would not be able to generate export orders with which to treat them.

Table F.1: Baseline Balance for Samples 1 and 2 (Appendix to Table 3)

	Sample 1			Sample 2		
	Control Group Mean	Treatment	Obs.	Control Group Mean	Treatment	Obs.
Panel A: Household Characteristics						
Age	51.4	-1.5	79	50.7	2.8	139
	(1.1)	(2.2)		(1.0)	(2.2)	
Number of years in rug business	39.3	-2.0	77	36.8	1.9	136
	(1.1)	(2.3)		(1.1)	(2.5)	
Illiterate?	0.70	0.14	79	0.59	0.07	135
	(0.05)	(0.10)		(0.04)	(0.10)	
Household size	4.5	-0.1	79	4.1	0.1	140
	(0.2)	(0.3)		(0.1)	(0.3)	
Digit Span Recall	5.0	-0.1	72	6.3	0.5	132
	(0.1)	(0.2)		(0.1)	(0.3)	
Cognitive Quiz	0.09	-0.03	79	0.06	0.08	140
	(0.03)	(0.06)		(0.03)	(0.09)	
Panel B: Firm Characteristics						
Price per square meter	35.6	6.6	79	25.4	2.1	139
	(5.9)	(11.9)		(1.8)	(3.9)	
Direct monthly profits from rug business	323	31.1	79	858	-12.5	139
	(57.8)	(117.0)		(38.6)	(85.3)	
Hours worked last month	208	-5.7	79	269	1.3	139
	(11.5)	(23.3)		(5.1)	(10.8)	
Number of employees	1.25	0.00	79	1.00		139
	(0.08)	(0.17)		-	-	
Total produced last month (m^2)	61.3	7.11	79	43.6	0.3	139
	(10.9)	(22.15)		(2.4)	(5.8)	
Ever exported?	0.03	0.00	79	0.17	0.03	140
	(0.02)	(0.04)		(0.03)	(0.08)	
Average Quality	0.22	-0.18 **	78	0.19	-0.09	140
	(0.05)	(0.08)		(0.03)	(0.06)	
Joint F-test		0.81			1.54	
Attrition in Follow Up Surveys	0.16	0.04	474	0.07	-0.04 *	560
	(0.02)	(0.04)		(0.01)	(0.02)	
Attrition in Quality Lab	0.08	0.18 **	79	0.16	-0.10 *	140
	(0.04)	(0.08)		(0.04)	(0.05)	

Notes: Table presents baseline balance for Sample 1 duble-firms (left panel) and the Sample 2 (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 and price 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 rows report average attrition across all survey rounds and the quality lab respectively. Significance * .10; ** .05; *** .01.

Table F.2: Impact of Intervention on Firms Knowingly Exporting: Sample 1 and Sample 2 (Appendix to Table 4)

	Sample 1		Sample 2		
	ITT	TOT	ITT	TOT	P-Value of TOT
	(1)	(2)	(3)	(4)	Comparison
Indicator for Ever Exported	0.31 **	0.79 ***	0.68 ***	0.75 ***	0.90
	(0.12)	(0.25)	(0.07)	(0.07)	
R-squared	0.12	0.34	0.45	0.49	
Control Group Mean	0.19	0.19	0.20	0.20	
Observations	59	59	132	132	

Notes: Table regresses an indicator for if a firm has ever knowingly produced rugs for export markets on indicators of treatment (column 1 & 3) or takeup (column 2 & 4). The question was asked in Round 5 for Sample 1 and Round 3 for Sample 2. The TOT regression instruments takeup with treatment. The last column reports the p-value from the statistical test of equivalence of the TOT coefficients between the two samples. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Significance * .10; ** .05; *** .01.

Table F.3: Impact of Exporting on Firm Profits: Sample 1 and Sample 2 (Appendix to Table 5)

Panel A: Monthly Profits

	Sample 1				Sample 2				
	ITT	TOT	Control Mean	Obs.	ITT	TOT	Control Mean	Obs.	P-Value of TOT Comparison
Outcome Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (Direct Profits)	0.27 **	0.83 **	890	198	0.25 ***	0.30 ***	951	375	0.11
	(.11)	(.33)			(.06)	(.07)			
Log (Reported Revenues - Reported Costs)	0.22 **	0.71 **	892	269	0.23 ***	0.28 ***	955	375	0.21
	(.11)	(.33)			(.05)	(.06)			
Log (Constructed Revenues - Constructed Costs)	0.14	0.48	940	310	0.24 ***	0.29 ***	957	375	0.59
	(.10)	(.34)			(.05)	(.07)			
Log (Hypothetical Profits)	0.39 **	1.29 **	465	314	0.36 ***	0.44 ***	586	373	0.19
	(.18)	(.64)			(.10)	(.12)			

Panel B: Monthly Profits per Owner Hour

	Sample 1				Sample 2				
	ITT	TOT	Control Mean	Obs.	ITT	TOT	Control Mean	Obs.	P-Value of TOT Comparison
Outcome Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Log (Direct Profits)	0.24 **	0.74 **	3.48	198	0.17 ***	0.21 ***	3.56	375	0.09
	(.10)	(.30)			(.05)	(.06)			
Log (Reported Revenues - Reported Costs)	0.21 **	0.66 **	3.49	262	0.15 ***	0.19 ***	3.58	375	0.12
	(.10)	(.30)			(.05)	(.06)			
Log (Constructed Revenues - Constructed Costs)	0.15 *	0.52 *	3.50	309	0.16 ***	0.19 ***	3.58	375	0.28
	(.09)	(.30)			(.05)	(.06)			
Log (Hypothetical Profits)	0.29 **	0.98 **	4.80	314	0.21 ***	0.26 ***	6.01	373	0.07
	(.11)	(.38)			(.07)	(.08)			

Notes: Table reports treatment effects on different real profit measures, all measured in logs seperately for Sample 1 duble-firms and Sample 2. See text for details regarding each measure. Panel A reports the analysis for the overall values whereas Panel B reports the results from the outcome per owner hour. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Control group means are reported in Egyptian pounds (LE) in Panel A and LE/hour in Panels B. P-Values from testing equivalence of the TOT's from each sample are reported in the last column. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table F.4: Sources of Changes to Firm Profits: Sample 1 and Sample 2 (Appendix to Table 6)

Panel A: Components of Profits

	Sample 1				Sample 2				
	ITT	TOT	Control Mean	Obs.	ITT	TOT	Control Mean	Obs.	P-Value of TOT Comparison
Outcome Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Output Prices	0.40 **	1.36 **	26.4	315	0.46 ***	0.56 ***	29.3	376	0.23
	(.18)	(.65)			(.10)	(.12)			
Output (m^2)	-0.30 *	-1.01 *	74.7	301	-0.22 **	-0.27 ***	57.9	375	0.21
	(.16)	(.59)			(.09)	(.10)			
Monthly Hours Worked	0.01	0.03	265.0	303	0.08 ***	0.10 ***	272.0	375	0.57
	(.03)	(.10)			(.02)	(.03)			
Number of Looms	0.01	0.04	1.2	318	-0.06 ***	-0.07 ***	1.1	376	0.62
	(.07)	(.22)			(.01)	(.02)			
Warp Thread Ball (kg)	-0.01	-0.03	5.2	311	0.13 **	0.15 **	6.3	377	0.13
	(.20)	(.68)			(.05)	(.06)			

Panel B: Inputs

	Sample 1				Sample 2				
	ITT	TOT	Control Mean	Obs.	ITT	TOT	Control Mean	Obs.	P-Value of TOT Comparison
Outcome Variable:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Weft Thread Price	0.16	0.62	12.3	188	0.23 ***	0.29 ***	13.1	376	0.54
	(.15)	(.53)			(.04)	(.05)			
Warp Thread Price	-0.03	-0.10	19.8	309	-0.03	-0.04	17.0	376	0.74
	(.05)	(.17)			(.03)	(.04)			
Weft Thread Quantity	-0.25	-0.85	130.0	302	-0.13	-0.16	97.8	375	0.26
	(.17)	(.60)			(.09)	(.10)			
Warp Thread Quantity	-0.01	-0.03	17.7	311	0.08	0.10	17.8	375	0.85
	(.20)	(.68)			(.09)	(.11)			

Notes: Table reports treatment effects on real prices, output, hours worked and size of the warp thread ball, all measured in logs, seperately for Sample 1 duble-firms and Sample 2. Monthly hours are calculated using average daily hours and number of days worked last month.The TOT regression instruments takeup with treatment. The TOT regression instruments takeup with treatment. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. P-Values from testing equivalence of the TOT's from each sample are reported in the last column. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table F.5: Impact of Exporting on Quality Levels: Sample 1 and Sample 2 (Appendix to Table 7)

Panel A: Quality Metrics

	Sample 1			Sample 2		
	Control Mean	ITT (1)	TOT (2)	Control Mean	ITT (3)	TOT (4)
Corners	2.89	0.66 ***	1.73 ***	3.00	1.38 ***	1.69 ***
		(0.21)	(0.41)		(0.13)	(0.08)
Waviness	2.89	0.67 ***	1.76 ***	3.02	1.36 ***	1.66 ***
		(0.20)	(0.36)		(0.13)	(0.08)
Weight	2.90	0.67 ***	1.75 ***	3.14	1.32 ***	1.60 ***
		(0.20)	(0.38)		(0.12)	(0.09)
Touch	3.14	0.24 ***	0.67 ***	3.11	0.54 ***	0.65 ***
		(0.08)	(0.18)		(0.08)	(0.06)
Packedness	3.17	0.39 **	1.32 ***	3.08	1.38 ***	1.68 ***
		(0.15)	(0.36)		(0.12)	(0.08)
Warp Thread Tightness	3.04	0.43 ***	1.38 ***	3.05	1.24 ***	1.51 ***
		(0.15)	(0.35)		(0.12)	(0.09)
Firmness	2.99	0.34 **	1.19 ***	2.97	1.43 ***	1.75 ***
		(0.14)	(0.35)		(0.13)	(0.08)
Design Accuracy	3.28	0.36 **	1.21 ***	3.12	1.22 ***	1.48 ***
		(0.15)	(0.35)		(0.12)	(0.10)
Warp Thread Packedness	2.90	0.66 ***	1.71 ***	3.10	1.33 ***	1.64 ***
		(0.20)	(0.40)		(0.13)	(0.09)
Inputs	3.09	0.44 ***	1.51 ***	3.06	1.37 ***	1.66 ***
		(0.15)	(0.38)		(0.11)	(0.09)
Loom	2.03	0.02	0.06	2.02	0.04	0.05
		(0.02)	(0.06)		(0.04)	(0.04)
R-squared		0.32	0.50		0.57	0.66
Observations		2,765	2,765		4,120	4,120

Panel B: Stacked Quality Metrics

	Sample 1			Sample 2		
	Control Mean	ITT (1)	TOT (2)	Control Mean	ITT (3)	TOT (4)
Stacked Quality Metrics	2.94	0.41 ***	1.26 ***	2.97	1.14 ***	1.39 ***
		(0.13)	(0.27)		(0.10)	(0.06)
P-Value of TOT Comparison						0.63
R-squared		0.29	0.44		0.52	0.60
Observations		2,765	2,765		4,120	4,120

Notes: The table reports the estimated impacts seperately for Sample 1 duble-firms and Sample 2. 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 F.6: Impact of Exporting on Productivity: Sample 1 and Sample 2 (Appendix to Table 8)

	Sample 1						Sample 2						
	ITT		TOT		Control Mean	Obs.	ITT		TOT		Control Mean	Obs.	P-Value of TOT Comparison
	(1)		(2)		(3)	(4)	(5)		(6)		(7)	(8)	(9)
Log(Output Per Hour)	-0.24		-0.79		0.30	311	-0.24	***	-0.29	***	0.24	376	0.47
	(0.15)		(0.53)				(0.09)		(0.10)				
Log(Unadjusted TFP)	-0.30	*	-1.00	*	0.58	299	-0.26	***	-0.32	***	0.43	375	0.32
	(0.15)		(0.55)				(0.09)		(0.10)				

Notes: Table reports treatment effects on the two productivity measures: (log) output per hour and (log) unadjusted TFP seperately for Sample 1 duble-firms and Sample 2. See Appendix A for the methodology used to obtain unadjusted 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; ***

Table F.7: Step 1: Quality and Productivity, Sample 1 and Sample 2 (Appendix to Table 9)

	Stacked Quality Metrics					Log(Output per Hour)					Log(TFP)				
	ITT	TOT	ITT	TOT		ITT	TOT	ITT	TOT		ITT	TOT	ITT	TOT	
Panel A: Specification Controls					P-Value of					P-Value of					P-Value of
	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT
	(1)	(2)	(3)	(4)	Comparison	(5)	(6)	(7)	(8)	Comparison	(9)	(10)	(11)	(12)	Comparison
Treatment	0.17 ***	0.72 ***	0.52 ***	0.82 ***	0.56	0.11	0.44	0.29 ***	0.45 ***	0.98	0.04	0.19	0.27 ***	0.42 ***	0.58
	(0.04)	(0.15)	(0.09)	(0.09)		(0.11)	(0.43)	(0.10)	(0.16)		(0.11)	(0.40)	(0.10)	(0.15)	
(log) Thread quantity	0.00	-0.03	0.02	0.04		-0.01	-0.03	-0.42 **	-0.40 **		0.08	0.06	-0.38 **	-0.36 **	
	(0.06)	(0.05)	(0.10)	(0.07)		(0.16)	(0.15)	(0.19)	(0.18)		(0.15)	(0.15)	(0.18)	(0.17)	
Difficulty Control	0.47 ***	0.33 ***	0.42 ***	0.33 ***		-0.17 **	-0.24 **	-0.11 **	-0.16 ***		-0.22 ***	-0.25 ***	-0.12 **	-0.17 ***	
	(0.03)	(0.05)	(0.04)	(0.04)		(0.07)	(0.10)	(0.05)	(0.06)		(0.07)	(0.09)	(0.05)	(0.05)	
(log) # colors	0.03 *	0.02	0.02	0.00		-0.07	-0.07 **	-0.07	-0.08 *		-0.08 **	-0.09 ***	-0.06	-0.07 *	
	(0.02)	(0.02)	(0.02)	(0.02)		(0.04)	(0.04)	(0.04)	(0.04)		(0.03)	(0.03)	(0.04)	(0.04)	
Low-market Segment	-0.24 ***	-0.07	-0.16 ***	-0.10 **		0.28 **	0.38 **	0.55 ***	0.59 ***		0.26 **	0.30 **	0.55 ***	0.58 ***	
	(0.05)	(0.06)	(0.04)	(0.04)		(0.12)	(0.17)	(0.10)	(0.10)		(0.11)	(0.15)	(0.09)	(0.10)	
Mid-Market Segment	-0.26 ***	-0.09	-0.11 **	-0.04		0.28 **	0.38 **	0.30 ***	0.34 ***		0.21 *	0.25 *	0.33 ***	0.36 ***	
	(0.06)	(0.07)	(0.05)	(0.05)		(0.12)	(0.15)	(0.10)	(0.10)		(0.11)	(0.13)	(0.09)	(0.10)	
Rug Type FEs	yes	yes	yes	yes		yes	yes	yes	yes		yes	yes	yes	yes	
Input Tread Type FEs	yes	yes	yes	yes		yes	yes	yes	yes		yes	yes	yes	yes	
R-squared	0.63	0.65	0.67	0.69		0.62	0.63	0.55	0.54		0.69	0.70	0.55	0.55	
Observations	2,744	2,744	4,076	4,076		302	302	371	371		290	290	370	370	
Panel B: Specification Fixed Effects															
	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT
	(1)	(2)	(3)	(4)	Comparison	(5)	(6)	(7)	(8)	Comparison	(9)	(10)	(11)	(12)	Comparison
Treatment	0.06	0.51	0.37 ***	0.71 ***	0.65	0.22 **	1.53	0.42 ***	0.81 **	0.47	0.24 **	1.48 **	0.38 ***	0.73 **	0.32
	(0.05)	(0.39)	(0.12)	(0.19)		(0.11)	(0.96)	(0.14)	(0.39)		(0.11)	(0.73)	(0.12)	(0.32)	
Specification FEs	yes	yes	yes	yes		yes	yes	yes	yes		yes	yes	yes	yes	
R-squared	0.71	0.71	0.73	0.74		0.77	0.70	0.64	0.61		0.78	0.77	0.65	0.64	
Observations	2,744	2,744	4,076	4,076		153	153	223	223		143	143	223	223	
Panel C: Specification-Adjusted Dependent Variables															
	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT	Sample 1		Sample 2		TOT
	(1)	(2)	(3)	(4)	Comparison	(5)	(6)	(7)	(8)	Comparison	(9)	(10)	(11)	(12)	Comparison
Treatment	0.21 ***	0.65 ***	0.60 ***	0.74 ***	0.47	0.04	0.14	0.34 ***	0.41 ***	0.00	0.09	0.31	0.42 ***	0.51 ***	0.00
	(0.06)	(0.11)	(0.06)	(0.04)		(0.12)	(0.39)	(0.07)	(0.09)		(0.12)	(0.36)	(0.07)	(0.08)	
R-squared	0.12	0.21	0.26	0.32		0.01	0.03	0.16	0.16		0.03	0.08	0.18	0.19	
Observations	2,784	2,784	4,076	4,076		307	307	371	371		301	301	370	370	

Notes: Table reports treatment effects on the stacked quality measures, and the two productivity measures seperately for Sample 1 duble-firms and Sample 2. See Appendix A for the methodology used to obtain unadjusted 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. For the TFP regressions, Panels A and B use the unadjusted TFP measure and Panel C using the adjusted TFP measure; see Appendix A for details. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table F.8: Step 2: Quality and Productivity on Identical Domestic Rugs, Sample 1 and Sample 2 (Appendix to Table 10)

Panel A: Quality Metrics

	Sample 1						Sample 2					
	Master Artisan			Professor			Master Artisan			Professor		
	Control Mean	ITT (1A)	TOT (1B)	Control Mean	ITT (2A)	TOT (2B)	Control Mean	ITT (3A)	TOT (3B)	Control Mean	ITT (4A)	TOT (4B)
Corners	3.27	0.47 *	1.29 **	3.41	0.14	0.42	3.22	0.89 ***	0.96 ***	3.27	0.40 **	0.42 **
		(0.25)	(0.56)		(0.20)	(0.52)		(0.16)	(0.15)		(0.17)	(0.18)
Waviness	3.24	0.30	0.88	3.27	0.19	0.60	3.14	0.72 ***	0.78 ***	3.33	0.29 *	0.30 *
		(0.25)	(0.56)		(0.19)	(0.47)		(0.17)	(0.16)		(0.15)	(0.16)
Weight	3.62	0.27	0.78	3.89	0.53 **	1.48 **	3.59	0.85 ***	0.96 ***	3.53	0.62 **	0.74 ***
		(0.20)	(0.50)		(0.22)	(0.74)		(0.16)	(0.16)		(0.24)	(0.26)
Packedness	3.32	0.31	0.84 *	3.31	0.07	0.24	3.28	1.09 ***	1.19 ***	3.26	0.43 ***	0.48 ***
		(0.22)	(0.49)		(0.19)	(0.50)		(0.15)	(0.15)		(0.14)	(0.15)
Touch	3.41	0.22	0.61	3.27	0.25	0.73	3.24	0.73 ***	0.80 ***	3.27	0.43 ***	0.47 ***
		(0.20)	(0.45)		(0.20)	(0.47)		(0.13)	(0.13)		(0.15)	(0.16)
Warp Thread Tightness	3.03	0.29 *	0.80 **	3.36	-0.02	-0.01	2.99	0.66 ***	0.71 ***	3.28	0.43 ***	0.49 ***
		(0.16)	(0.35)		(0.19)	(0.53)		(0.10)	(0.10)		(0.15)	(0.15)
Firmness	3.32	0.22	0.55	3.24	0.074	0.24	3.16	1.04 ***	1.13 ***	3.22	0.44 ***	0.49 ***
		(0.26)	(0.65)		(0.19)	(0.50)		(0.14)	(0.14)		(0.15)	(0.16)
Design Accuracy	3.75	0.30	0.81 *	3.55	-0.01	-0.06	3.61	0.68 ***	0.79 ***	3.41	0.45 ***	0.48 ***
		(0.18)	(0.46)		(0.21)	(0.60)		(0.14)	(0.15)		(0.12)	(0.14)
Warp Thread Packedness	3.16	0.51 **	1.40 **	3.28	0.14	0.44	3.00	1.12 ***	1.20 ***	3.17	0.57 ***	0.65 ***
		(0.25)	(0.58)		(0.19)	(0.51)		(0.16)	(0.16)		(0.15)	(0.16)
R-squared		0.11	0.29		0.16	0.14		0.31	0.34		0.10	0.08
Observations		593	593		589	589		1,087	1,087		1,078	1,078

Panel B: Stacked Quality Metrics

	Sample 1						Sample 2					
	Master Artisan			Professor			Master Artisan			Professor		
	Control Mean	ITT (1A)	TOT (1B)	Control Mean	ITT (2A)	TOT (2B)	Control Mean	ITT (3A)	TOT (3B)	Control Mean	ITT (4A)	TOT (4B)
Stacked Quality Metric	3.35	0.32 *	0.91 **	3.40	0.15	0.43	3.25	0.87 ***	0.95 ***	3.30	0.45 ***	0.50 ***
		(0.18)	(0.40)		(0.16)	(0.40)		(0.12)	(0.11)		(0.12)	(0.13)
P-Value of TOT Comparison									0.92			0.88
R-squared		0.09	0.30		0.14	0.22		0.29	0.34		0.09	0.09
Observations		593	593		589	589		1,087	1,087		1,078	1,078

Panel C: Additional Quality Metrics

	Sample 1			Sample 2		
	Control Mean	ITT (1A)	TOT (1B)	Control Mean	ITT (3A)	TOT (3B)
Length Accuracy	-4.41	0.70	1.93		1.93 ***	1.95 ***
		(0.85)	(2.22)		(0.63)	(0.72)
Width Accuracy	-2.22	-0.21	-0.46		0.43	0.47
		(0.52)	(1.50)		(0.34)	(0.38)
Weight Accuracy	-197.0	83.0 ***	236.0 **		93.3 ***	108.0 ***
		(26.8)	(94.4)		(29.1)	(30.6)
Time (in minutes)	255.0	-22.10 *	-63.00 *		5.52	5.4
		(12.1)	(35.7)		(7.4)	(7.9)
R-squared		0.87	0.80		0.83	0.83
Observations		264	264		484	484

Notes: Table reports ITT and TOT specifications using the 9 quality metrics from the quality lab seperately for Sample 1 duble-firms and Sample 2. 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 objective accuracy measures which are calculated as the negative of the abosulte error for the specification, so that 0 is perfect and a higher value is better. It also includes 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 F.9: Step 3: Treatment Dynamics, Sample 1 and Sample 2

	Log(Direct Profits)		Log(Direct Profits per Hour)		Log(Output per Hour)	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)
Treatment	0.36	0.24 ***	0.31	0.20 ***	-0.24	-0.28 **
	(0.21)	(0.08)	(0.19)	(0.07)	(0.22)	(0.12)
x 2nd Round Since Treatment	-0.23	0.05	-0.17	-0.01	0.23	0.11
	(0.28)	(0.08)	(0.25)	(0.08)	(0.21)	(0.14)
x 3rd Round Since Treatment	-0.17	-0.03	-0.21	-0.09	-0.14	0.01
	(0.24)	(0.09)	(0.23)	(0.09)	(0.27)	(0.13)
x 4th Round Since Treatment	-0.16		-0.12		-0.28	
	(0.25)		(0.23)		(0.26)	
x 5th Round Since Treatment	0.16		0.16		0.14	
	(0.23)		(0.21)		(0.29)	
R-squared	0.15	0.29	0.15	0.18	0.18	0.19
Observations	198	375	198	375	311	376

	Log (Unadjusted TFP)		Log (Specification Adjusted TFP)		Log (Specification Adjusted Quality)	
	Sample 1	Sample 2	Sample 1	Sample 2	Sample 1	Sample 2
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)
Treatment	-0.31	-0.28 **	0.25 **	0.41 ***	-0.03	0.63 ***
	(0.20)	(0.11)	(0.10)	(0.09)	(0.05)	(0.08)
x 2nd Round Since Treatment	0.33	0.08	0.08	-0.03	0.16	-0.03
	(0.24)	(0.13)	(0.11)	(0.11)	(0.10)	(0.08)
x 3rd Round Since Treatment	-0.17	-0.02	-0.01	0.04	0.34 ***	-0.06
	(0.24)	(0.12)	(0.11)	(0.12)	(0.11)	(0.08)
x 4th Round Since Treatment	-0.29		-0.29		0.38 ***	
	(0.25)		(0.19)		(0.10)	
x 5th Round Since Treatment	0.15		0.00		0.27 ***	
	(0.24)		(0.15)		(0.09)	
R-squared	0.31	0.18	0.09	0.18	0.13	0.26
Observations	299	375	671	370	2,784	4,076

Notes: Table reports treatment effects interacted with dummies for each round of data collection separately for Sample 1 duble-firms and Sample 2. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Control group means are reported in levels. Standard errors are clustered by firm. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table F.10: Step 4: Information Flows and Quality Levels: Sample 1 and Sample 2 (Appendix to Table 11)

	Stacked Quality Metrics			Specification Adjusted Stacked Quality Metrics		
	Sample 1	Sample 2	P-Value of TOT	Sample 1	Sample 2	P-Value of TOT
	(1)	(2)	Comparison	(3)	(4)	Comparison
$Takeup_i$ x {Talked About Dimension}$_{id}$	0.29 ***	0.13 *	0.08	0.23 ***	0.10 *	0.17
	(0.07)	(0.07)		(0.08)	(0.05)	
Quality Metric FEs	yes	yes		yes	yes	
$Takeup_i$ x Quality Metric FEs	yes	yes		yes	yes	
Product characteristic controls	no	no		yes	yes	
Specification-adjusted Quality Metrics	no	no		no	no	
R-squared	0.79	0.82		0.50	0.59	
Observations	602	1,098		599	1,068	

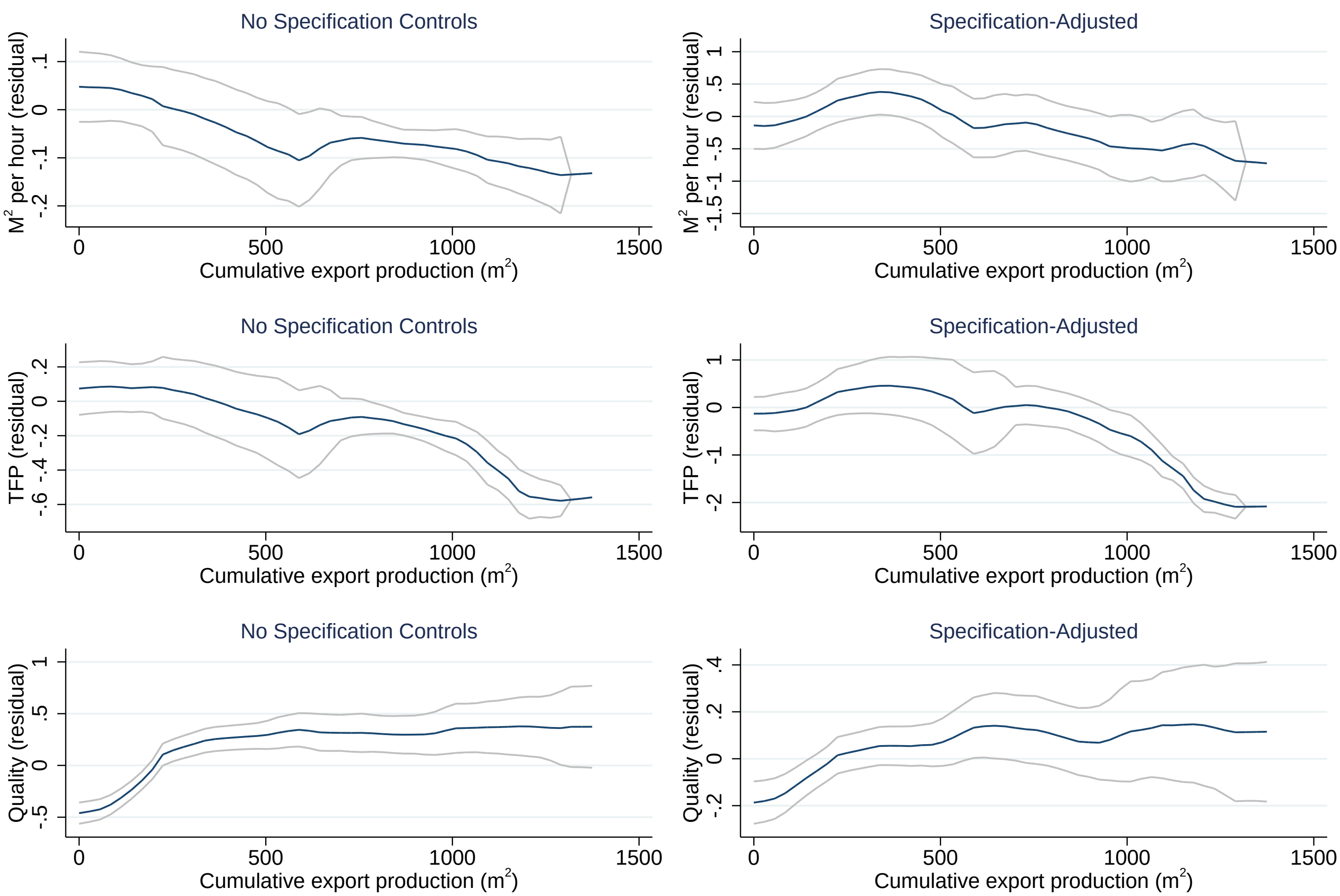
Notes: Table regresses stacked quality metrics on on takeup indicator and its interaction with a dummy that takes the value 1 if the intermediary talked to the firm about the particular quality metric. Table reports these results seperately for Sample 1 duble-firms and Sample 2. Columns 3-4 control for 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, quality metric fixed effects, and quality metric fixed effects interacted with takeup. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table F.11: Results for Non-Duble Strata

	ITT (1)	TOT (2)	R-squared (3)	Obs. (4)
Ever Exported	0.03	0.22	0.28	90
	(.09)	(.63)		
Direct Log Monthly Profits	-0.01	-0.07	0.05	324
	(.08)	(.76)		
Direct Log Profits per Hour	0.00	-0.02	0.04	323
	(.07)	(.72)		
Log Output per Hour	0.06	0.73	0.71	384
	(.10)	(1.30)		
Log Stacked Quality	-0.02	-0.28	0.56	2283
	(.04)	(.55)		
Log Unadjusted TFP	0.08	0.95	0.64	371
	(.12)	(1.43)		
Log Adjusted TFP	-0.04	-0.59	0.31	345
	(.09)	(1.27)		

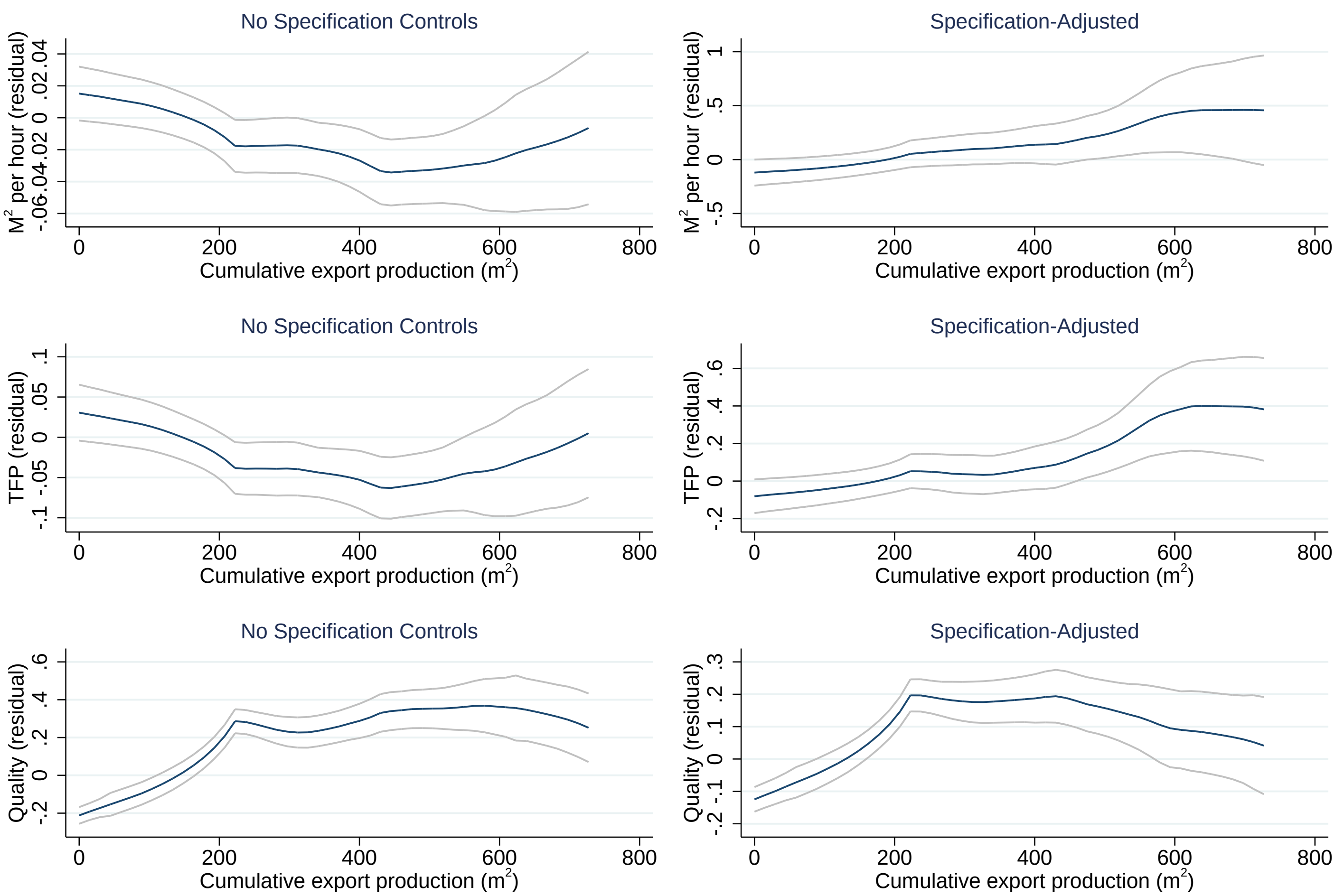
Notes: Table reports treatment effects on 8 main outcomes for the firms in our sample who were not in the duble strata. Due to the inability to secure export orders for non-duble rug types there is no first stage and no subsequent impacts. We were not able to secure funding to continue to collect longer term data on the non-duble strata, which is why the number of observations are low. We only collected the "ever export" variable on the "tups" stratum, which had the highest take up out of all orther strata. Regressions include baseline variables and round and strata fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Figure F.1: Step 3: Learning Curves, Sample 1 Takeup Firms (Appendix to Figure 5)



Notes: Figure plots learning curves obtained by regressing quality or productivity measures on firm and round fixed effects and then plotting a kernel-weighted local polynomial of the residuals against cumulative export production for takeup firms in Sample 1.

Figure F.2: Step 3: Learning Curves, Sample 2 Takeup Firms (Appendix to Figure 5)



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