Exporting and Firm Performance: Evidence from a Randomized Trial*

David Atkin† Amit K. Khandelwal‡ Adam Osman§

First Draft: June 2014
This Draft: October 2014

Abstract

We conduct a randomized control trial that generates exogenous variation in the access to foreign markets for rug producers in Egypt. Using this methodology and detailed survey data, we causally identify the impact of exporting on firm performance. Treatment firms report 15-25 percent higher profits and exhibit large improvements in quality alongside reductions in quantity-based productivity 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 induces changes in technical efficiency. First, treatment firms have higher productivity and quality after accounting for rug specifications. Second, when asked to produce an *identical* domestic rug using the same technology, treatment firms receive higher quality assessments despite no difference in production time. Third, treatment firms exhibit learning curves over time for both quality and productivity. Finally, we document knowledge transfers between buyers, intermediaries and producers 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, David McKenzie, Nina Pavcnik, Jim Tybout, 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 Chazen Institute at Columbia Business School. Yale IRB Approval #1001006247, Columbia IRB Approval #AAAE9678.

†UCLA, *email*: atkin@econ.ucla.edu

‡Columbia Business School, *email*: ak2796@columbia.edu

§University of Illinois at Urbana-Champaign, *email*: aosman@illinois.edu

1 Introduction

Recent decades have seen large resources flow into "Aid-for-Trade" and market access initiatives in developing countries. For example, the WTO Aid-for-Trade Initiative has secured $48 billion in annual commitments from donors to help developing countries overcome "trade-related constraints". The aim of these interventions is to bring about growth and reduce poverty through improvements in firm performance. Central to this goal is the concept of learning-by-exporting, improvements in technical efficiency induced by exporting (Clerides et al., 1998, de Loecker, 2007). Such learning processes are required to generate growth beyond the gains generated by static trade models (Grossman and Helpman, 1993, Harrison and Rodriguez-Clare, 2010).

Despite their pervasiveness, we know very little about the efficacy of these policy initiatives in improving firm performance, and if they are effective, whether these improvements occur through learning-by-exporting or other mechanisms (Fernandes et al., 2011). There are two central challenges in identifying potential causal effects of exporting. First, more productive firms select into exporting (Bernard and Jensen, 1999, Melitz, 2003). This selection has plagued attempts to identify empirically the learning-by-exporting mechanism because what appears to be higher productivity among exporters may simply be self-selection into export markets. The second difficulty is that researchers typically lack detailed information required to isolate changes that occur within firms due to exporting. Instead, the literature typically uses residual-based measures, such as total factor productivity, which ultimately may capture a multitude of non-learning mechanisms such as changes in product specifications, markups, or input costs (de Loecker and Goldberg, 2014).

This paper conducts a randomized field experiment on rug manufacturers in Egypt to examine the channels through which exporting affects the performance of firms. To our knowledge, this is the first attempt to generate exogenous firm-level variation in the opportunity to export. The random assignment into exporting directly addresses the first challenge: selection of firms into exporting. Specifically, we provided a subset of small rug producers 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 these orders by producing 110 square meters (m^2) of rugs (approximately eleven weeks of work). As in any 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 small number of firms had ever knowingly exported their products. Hence, we interpret our experimental design as providing non-exporting firms with the opportunity to export to high-income markets.

To address the second challenge in identifying the impact of exporting, we tracked performance measures through periodic surveys of both treatment firms (who received the opportunity to export) and control firms (who received no such opportunity). Focusing the analysis on a single industry, and specifically handmade rugs, provides several advantages; the production technology is homogenous across firms, quality metrics are well-defined and codifiable, and physical productivity can be accurately measured. For example, the literature typically relies on prices and input costs to infer product quality (e.g., Schott, 2004 or Hallak, 2006). In contrast, our production-line level data allow us to record detailed product specifications for the rugs being produced at the time of each survey round. These are attributes associated with quality, such as the thread count, that a buyer chooses when they place their order. We complement these specification measures with direct measures of product quality along 11 dimensions from a skilled quality assessor who visited each firm in each survey round. These quality measures capture a combination of both specifications and hard-to-codify characteristics that depend on the skill of the firm such as how flat the rug lies on the floor. Additionally, for our treatment firms we have high-frequency data from the intermediary's order book that records quality metrics for every rug these firms export. We also collected information flows between buyers, intermediaries and producers that include transcripts of buyer feedback and the content of discussions between the intermediary and the producers. Together, these data allow us to address directly the measurement challenges in assessing how exporting affects firm performance.

Thanks to the randomization procedure, the analysis of the intervention is straightforward: the causal effects of exporting are identified by comparing mean outcomes between treatment and control firms. We find that the opportunity to export raises the overall performance of firms as measured by profits. Treatment firms report 15-25 percent higher profits relative to control firms, depending on the profit measure. The substantial increase in profits is interesting in itself, particularly given the more moderate profit impacts the literature has found when exploring supply-side interventions such as credit access (Banerjee, 2013), and are suggestive that the distributional consequences of trade may come in part from heterogeneity in market access. While it is important to understand whether the market access program is cost effective and/or alleviates market failures, both questions we plan to address in future work, this paper focuses on understanding precisely how improved market access affects the performance of firms.

Guided by a simple theoretical framework, the remainder of the paper explores the sources of the rise in profits. Treatment firms increase scale, measured by increases in total labor hours and through longer production runs. Treatment firms also report increases in output and input prices. However, despite the price increases, we observe a *decline* in total output (m^2 of rugs produced) among treatment firms. This seemingly puzzling finding suggests that the impact of exporting in our setting does not work solely through scale or price effects. Our data confirm that the opportunity to export significantly raises quality levels along virtually every quality dimension. At the same time, quantity-based productivity (not adjusted for rug specifications) falls by 24 percent among treatment firms.[1] These findings are consistent with the fact that buyers from high-income countries demand higher-quality rugs that are more difficult to make and hence slower to produce.

[1] If we inferred productivity from revenues, we would conclude that productivity increases. This result underscores the importance of separately observing quantities and prices, a point emphasized by de Loecker (2011).

Quality upgrading can arise through two distinct channels which we term *passive-quality-upgrading* and *learning-by-exporting*. In a passive-quality-upgrading story, firms already know how to manufacture high-quality products. The export opportunity exposes firms to buyers who demand high-quality rugs, and so firms will raise the specifications (and hence the quality) of the rug as long as it is profitable to do so. In contrast, learning-by-exporting occurs when there are changes in technical efficiency induced by exporting. Increases in technical efficiency can occur either by producing more output per input (for a given set of product specifications) or by producing higher quality conditional on specifications. If these increases in efficiency are biased towards the production of high-quality rugs, both rug quality and profits will rise. This mechanism is fundamentally distinct from a passive-quality-upgrading story where quality can rise without any changes in technical efficiency. Unlike previous studies, we can distinguish between these two modes of upgrading because of the experimental variation and because we collect direct information on productivity, rug quality and rug specifications.

We use four pieces of evidence to detect learning-by-exporting. The first is that both quality and productivity rise after conditioning on product specifications (recall that without conditioning, productivity falls). In a pure passive-quality-upgrading story, since the parameters of the production function are not changing, quality and productivity should remain constant once we adjust for product specifications. Second, at the endline, we asked all firms in our sample to manufacture an *identical* domestic rug using a loom in a workshop we leased. Treatment firms produced higher quality rugs along every quality metric as well on objective measures of size and weight accuracy; moreover, treatment firms *did not* take longer to produce these rugs despite their higher levels of quality. The third piece of evidence comes from exploring the evolution of quality and productivity over time. Inconsistent with a passive-quality-upgrading story where quality should immediately jump and then stay fixed, we find strong evidence of a quality learning curve: rug quality increases with cumulative export production. Similarly (unadjusted) productivity initially drops upon exporting and then gradually rises over time (while adjusted productivity simply increases over time). Finally, we draw on correspondences between foreign buyers and intermediaries, 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 under a passive-quality-upgrading story). 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.

While expanding access to domestic markets may also generate profit increases, it is unlikely to generate the quality upgrading and associated learning we find since there are only a limited number of high-income/sophisticated consumers in a developing country such as Egypt (see Park et al., 2010 and Artopoulos et al., 2013 for further discussion of the particular importance of export market access for firms in developing countries).[2]

[2]Exporting has further distinct effects beyond generic demand expansions by protecting firms from demand volatility in the domestic market through access to multiple markets. For example, as would be predicted by a trade model with country-specific demand shocks, we find that profit volatility is significantly lower among treatment firms.

With regards to external validity, there are both advantages and disadvantages to our setting. The process of exporting via an intermediary is common to firms in many industries and countries and handmade home furnishings is a large industry in its own right (as we discuss in the next section).[3] The firms in our sample are small, typically having only one employee, and production is not automated. However, our results leave open the possibility that the learning-by-exporting mechanism is stronger in settings with larger scope for upgrading. (And of course, it is precisely their small size that allows us to assemble a large sample necessary for inference in the first place.) Ultimately, the external validity of our results is an empirical question, and the novel methodology we propose in this paper can be applied to other industries or populations to further our understanding of the impacts of exporting on firm performance.

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.[4] The evidence from these studies is mixed. For example, de Loecker (2007) uses matching techniques and finds evidence supporting learning-by-exporting as do Park et al. (2010) exploiting exchange rate shocks. Clerides et al. (1998) finds no evidence of breaks in firms' cost curves upon export market entry, suggesting no learning is present, while Aw et al. (2011) find evidence of learning-by-exporting among Taiwanese firms. Our study contributes to this debate by directly confronting issues of selection and measurement that often clouds this literature.

Given the finding of substantial increases in quality among exporters, the paper is also closely linked 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). A more recent series of firm-level studies suggest that quality upgrading is paramount for export success.[5] Unlike much of this literature that must infer quality from price data or international certifications, or through structural models where quality is inferred from prices and quantities, we collect direct measures of quality.[6] In addition to our randomization methodology and the comparatively rich survey data, we contribute to this literature by carefully distinguishing between passive-quality-upgrading and quality upgrading that occurs through learning-by-exporting.

Finally, although the use of randomized control trials is novel in the trade literature, the methodology has been used to understand supply constraints in firms (e.g., de Mel et al., 2008, de Mel et al., 2010, Bloom et al., 2013 and de Mel et al., 2014 explore credit constraints, input market frictions and managerial constraints). We complement this literature by providing the first

[3]World Bank Enterprise surveys report that 36 percent of exporting firms use an intermediary (62 percent for firms with 5 or fewer employees). Exporting in this manner may be particularly prevalent in the rug industry. For example, Chinese customs data show that 37 percent of Chinese exports in HS Code 580500 ("hand-woven tapestries of the type Gobelins, Flanders, Aubusson, Beauvais and the like, and needle-worked tapestries") went through intermediaries compared to 20 percent of overall exports.

[4]This literature includes Clerides et al. (1998), Bernard and Jensen (1999), Van Biesebroeck (2005), de Loecker (2007, 2013), Park et al. (2010) and Aw et al. (2011). For surveys of the literature, see Keller (2004), and Wagner (2007).

[5]See Verhoogen (2008), Crozet et al. (2012), Brambilla et al. (2012) and Hallak and Sivadasan (2013).

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

experimental evidence for the importance of demand constraints and the effects of relaxing those constraints through a market access initiative.

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 caused by the intervention and Section 5 decomposes the profit changes. Section 6 first lays out a simple theoretical framework that then guides our approach to detecting learning-by-exporting. Section 7 concludes.

2 Research Setting

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

2.1 The Industry and the Location

In order to carry out a randomized evaluation of the impact of exporting, we sought out governmental and non-governmental organizations involved in market access initiatives. In October 2009, we entered conversations with Aid to Artisans (ATA), a U.S.-based NGO with a mission to create economic opportunities for small-scale producers of handmade products around the world. They had recently acquired USAID funding for a market access facilitation program in Egypt and we agreed to work with them to evaluate the program.

ATA's program in Egypt followed their standard protocol for generating successful exporting relationships between small-scale developing-country producers and high-income OECD markets. First, ATA explores the country in question for products that would both appeal to high-income OECD consumers and be priced competitively. Once candidate products are found, ATA identifies a lead intermediary based in the developing country. The lead intermediary assists in finding small-scale producers that can manufacture the products, is the conduit for passing information and orders between the producers and the buyers, and handles the export logistics required to ship the products to importers or retailers abroad. ATA provides training to the intermediary and then works closely with it to both produce appealing products and to market them. To produce appealing products, ATA draws on its experience in the handcrafts industry and will occasionally pay for design consultants. In terms of marketing the products, ATA prominently displays the products at major trade shows, for example the biannual New York International Gift Fair (NYIGF) which draws 35,000 attendees each year, as well as drawing on its extensive network of contacts in the industry.

Working through a lead intermediary firm, rather than matching individual producers directly with foreign buyers, is an important aspect of ATA's business model. The ultimate objective for ATA is to foster self-sustaining relationships whereby it can eventually exit the sector. The presence of a lead intermediary makes this possible as it would be too costly for buyers to contract with many small producers in the absence of ATA. The hope is that with ATA's training, the intermediary develops the necessary skills to maintain and even expand its relationships with clients once ATA departs.[7]

This process through which exporting relationships emerge is not uncommon in other settings. Ahn et al. (2011) show that small-scale firms are likely to use intermediaries to export in order to avoid large fixed costs associated with directly exporting. World Bank Enterprise Data record direct and indirect export activity of firms across many countries. Among manufacturing firms, 36 percent of exporters use an intermediary with this number rising to 62 percent when we restrict attention to firms with five or fewer employees to facilitate comparison with our context. In our setting, the lead intermediary fulfills the role of aggregating orders and spreading the fixed costs of exporting across many small-scale producers. ATA acts as another middleman in this process, facilitating connections between domestic intermediaries in developing countries and foreign buyers.[8]

Alongside ATA, we searched for viable Egyptian products for more than a year before identifying handmade carpets from Fowa as having potential. Fowa is a peri-urban town located two hours southeast of Alexandria. The town has a population of 65,000 and lies in the governorate of Kafr El-Sheikh which has an average income per capita of $3,600 (PPP-adjusted), well below the national average of $6,500 (PPP-adjusted). Fowa is well known for its carpet cluster which contains hundreds of small firms that manufacture handmade textile products using wooden looms. These firms, typically employing between 1 to 4 employees, predominantly produce flat-weave rugs, a product in which Egypt has a strong historical reputation.

Both the handmade craft industry and the carpets and 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, 2010). Egypt is the 11th largest producer of carpets and rugs with a total production at $734 million; this represents 36 percent of Egypt's total textile sector and alone accounts for 1.3 percent of total manufacturing output.[9] More than 17,000 people were engaged in the carpets and rugs industry in Egypt, representing nearly 7 percent of world employment in this industry and 1.7 percent of total manufacturing employment in Egypt (UNIDO, 2013). According to official trade statistics, in 2013 Egypt's exports in HS58 (special woven fabrics, tufted textiles, lace) constituted 0.6 percent percent of its total exports which exceeded Egypt's share of total world exports (0.2 percent) indicating that Egypt has a revealed comparative advantage in this sector. The U.S. is the largest importer of Egypt's exports in HS58 accounting for 38 percent of Egypt's sales while advanced economies in Europe account for 31 percent.

In November 2010 we identified a local carpet intermediary, Hamis Carpets, as a potential partner for the program. Hamis is the largest intermediary in Fowa and accounts for around

[7] For example, at the 2010 NYIGF, ATA introduced us to several intermediaries from developing countries who had initially been linked to Western markets through ATA and had now graduated to having their own independent display at the gift fair.

[8] The need for such links between intermediaries/suppliers at the source and buyers at the destination has also been noted by Rauch (1999), Rauch and Trindade (2002) and Feenstra and Hanson (2004).

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

20 percent of the market. At the time, Hamis Carpets earned 70 percent of its sales in the domestic Egyptian market, mostly selling to distributors and retailers in Cairo and other Egyptian tourist markets such as Alexandria and Luxor, but expressed an interest in expanding its overseas sales. ATA believed that, together with Hamis, they could generate additional orders from overseas buyers and fill these order by forming new relationships with small-scale producers in Fowa. ATA brought the CEO of Hamis to the US for a training course, provided marketing support and insisted that Hamis agree to the protocols of our experiment, which we describe below. The marketing support included displaying the new products at the NYIGF, the Atlanta International Gift & Home Furnishings Market, the Paris Maison & Object Trade Fair, the International Handicrafts Trade Fair in Florence, Italy as well as at smaller events in Cairo. ATA also arranged for Hamis' rugs to be marketed to high-income OECD retailers by a US-based rug importer.[10]

There are both advantages and disadvantages to our setting. In terms of advantages, first, although the products are handmade, the process of exporting via an intermediary is similar to how other industries are organized, as described above. Second, in part because the producers are small in size, we were able to locate a large number of independent producers in a close geographic proximity making it possible (and economically feasible) to run a randomized control trial. Third, the production technology (described below) is identical across all the firms which facilitates data collection (we can tailor the surveys to ask specific questions about production) and makes it easy to compare outcomes across producers.[11] It also makes it relatively easier to identify heterogeneous impacts of the intervention because all firms are *ex ante* very similar, a point that has been emphasized by McKenzie (2012) and McKenzie and Woodruff (2013). Fourth, handmade home furnishings is an important industry, particularly when considering sectors in which exporting can reduce poverty. As mentioned above, there is a sizable global market for handmade products to furnish homes, and it is one of the few industries in which small-scale producers located in remote regions of the developing world can feasibly export their products to buyers in developed countries. At the same time, many handicraft workers are poor. Accordingly, policymakers are interested in knowing whether encouraging these industries can reduce poverty: Egypt's Minister of Industry and Foreign Trade recently stated, "Handicrafts are considered one of [the economy's] vital industry sectors...The sector's major importance comes as it is a heavily labor-intensive and a major employer, it also does not require large funds to operate."[12]

In terms of disadvantages, the downside to having many firms is that the firms are small,

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

[11] For example, if firms were manufacturing both carpets and towels, it would be harder to compare dimensions of quality across products.

[12] http://www.dailynewsegypt.com/2013/12/15/foreign-trade-minister-issues-decision-to-establish-handicrafts-export-council

typically having only one employee. Exports are unlikely to have transformative effects on these firms. Likewise, the nature of the production technology is unique and the scope for technological upgrading is more limited that in modern industries that involve substantial automation. However, this suggests that any export-induced learning mechanisms that we observe in the data may also be present in other settings with larger scopes for upgrading.

2.2 Production Technology

The producers of handmade carpets in Fowa are firms that typically consist of a single owner who operates out of his (all producers in our sample are men) home, or a rental space. 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 rug types and are shown at the top of Figure 1. Kasaees rugs, the bottom-left rug in Figure 1, are lower-cost rugs woven from rags that are often used outside a house, for example as a door mat. Goublan rugs, depicted in the bottom-right of Figure 1, are the most expensive rug type, are typically used as wall hangings and are considered works of art. Producers of goublan rugs are widely believed to be the best weavers as these rugs involve very intricate designs; the local market perceives these rugs to be of the highest quality relative to the other three rug types. Duble rugs are the main rug type that Hamis Carpets, our intermediary, sells. As we explain later, our export orders ended up being almost exclusively for duble-type rugs.

The process of producing rugs is standardized across firms. The elements of the production technology are marked in Figure 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 number of openings per meter in the reed determines how finely woven the rug can be; more openings per meter allows for more intricate designs. The *weft thread* (typically made from wool) is weaved between these warp threads using a shuttle that the weaver maneuvers back and forth across the width of the rug. A heddle is used, in combination with the reed, to raise every alternate warp thread into a higher position allowing the weaver to quickly weave the weft threads between the warp threads. A weaver typically begins the weaving process by installing the warp thread on the base of the loom and runs the warp thread through the reed and heddle. He then uses the foot-operated heddle to push the warp thread up and down as he weaves the weft thread through horizontally. He continues to do this, changing out the weft thread based on the needs of the design, until he completes the rug. At that point he cuts off the completed rug and continues to utilize the remaining warp thread until he completes his production run of that particular type of rug.

The average duble rug is sold by the firms for LE42.5 per m^2, while tups and goublan rugs are about three times more expensive and kasaees rugs are about one-fifth the price of a duble rug.[13]

[13]The exchange rate on December 1, 2010 was 5.75 Egyptian pounds (LE) to 1 U.S. dollar. The exchange rate on July 1, 2014 was LE 7.15 to 1 U.S. dollar. We will apply an average exchange rate of 6.45 where we convert to U.S. dollars.

These price differences map to the large variation in hours required to produce a m^2 of rug. A typical duble rug requires 5.9 hours per m^2, while tups and goublan rugs require 6-8 times more hours per m^2 and kasaees rugs require one-fifth of this time. Irrespective of the rug type, after accounting for input costs and labor hours, hourly wages are roughly LE3.

There are several dimensions through which the quality can vary across firms within rug types. Firms can manufacture the rugs with different qualities of input thread which maps to higher output prices. As such, prices convey some information about product quality, but are not the only sources of variation. How well the warp and weft thread are installed on the loom also affects the extent to which a rug lies flat on a hard surface. The weaving technique also has large effects on the rug's quality. For example, poor weaving technique will effect the packedness of the rug. Packedness refers to how well the rug holds itself together, and lower quality rugs are either too packed, not packed enough, or have an inconsistent level of packedness throughout the rug. Additional measures of rug quality include whether it adheres to the desired size specifications, whether the rug is the correct weight, how well defined the corners and edges of the rug are, and how the rug feels to the touch. Since the rugs are made by hand, adhering to these quality attributes requires both effort and skill. As discussed below, we have collected these quality metrics for all firms.

2.3 Generating Export Orders

It took the combination of ATA and Hamis Carpets more than two years to generate sustained export orders from high-income OECD clients. The handmade textile market is a very competitive industry and it typically takes several attempts to identify products that can be successfully imported and marketed in high-income OECD markets. Conversations with ATA's staff revealed that only 1 in 7 potential matches ultimately leads to a sustained exporting relationship; most matches never move beyond trial orders. This is consistent with Eaton et al. (2013) who estimate that only 1 in 5 potential importer-exporter matches results in a business relationship.

While ATA was involved, they would typically introduce Hamis to foreign importers or retailers. Hamis would elicit preferences of the retailer in terms of price point, rug size, design and time frame. Foreign buyers typically provide Hamis with the detailed design and color patterns; as an example, see Figure 3. Hamis would organize the production of sample orders, either from its in-house weavers or ordered from one of the treatment firms in our sample.[14] This process can be costly, particularly in terms of time, as Hamis and the potential buyer iterate constantly on design patterns, color schemes, technical aspects of quality, and price. For example, one foreign importer who was marketing Hamis' rugs to a German retailer complained about a large sample order in an email to Hamis:

> ...[Name redacted] did not accept the finishing of the collections. The labels were wrong, colours were in wrong chronology, the lugs were dented.[15] So we got no money from them and they did not take the samples. We are afraid they will cancel all business

[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] A lug is a circular piece of metal placed in the corner of sample rugs to make them easier to display.

> with us because of that. We have an outstanding balance with this company of nearly €10,000. This is the truth and you have part of this disaster.

Hamis responded to this complaint noting:

> It isn't acceptable to give such a reply because you asked for an order and we did it and now you refuse to receive it. How can we afford this loss? Regarding the mistake in putting color signs on samples, it isn't our fault because you changed the names and numbers more than once to cause this mess!

Such back-and-forth exchanges illustrate the challenges of inherently subjective aspects of production and the language barriers present (both Hamis and the client quoted above communicate in English, which is not the native language of employees in either firm).

The majority of rugs that have been demanded by foreign buyers are duble rugs, although one client ordered kasaees rugs. There have been no orders for goublan rugs, even though the local market in Egypt perceives these rugs to require the most skilled weaving techniques. But as Figure 1 illustrates, the style of goublan rugs is unlikely to appeal to high-income OECD buyers. Instead, it appears that high-income OECD buyers prefer "modern" designs, as illustrated in Figure 4. (The right-most rug in this figure is produced by one of our sample firms and retails for $1,400 in a high-end furniture store in the United States.)

After two-and-a-half years of searching, in June 2012, Hamis Carpets secured its first large export order (3,640 square meters) from a German buyer and since then, has generated additional large export orders. As of June 2014, its major buyers continue to place large, regular orders from Hamis. Figure 5 shows that cumulative export production since December 2010 have totaled 33,227m^2. Our records indicate that cumulative payments to the producers have totaled LE982,351 ($152,302). 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

This section describes the sample, experimental design, the specifics of the treatment intervention and measurement of the key variables.

3.1 Experimental Design

In July 2011, we compiled a listing of firms who worked on their own account, meaning that they bought their own inputs, had fewer than 5 employees, and had never previously worked with Hamis Carpets.[16] We relied on the help of an Egypt-based NGO to locate these firms since there is no census of carpet manufacturers in Fowa (all the firms in our sample are informal) and many firms are located within homes making them particularly difficult to find. These firms primarily specialize in one of the four rug types described above, and we stratified the sample both on the type of rug they produced and the loom size. We stratified on rug type because of the possibility that ATA and Hamis would not secure export orders for each rug type, which turned out to be the case. We stratified on loom size because the loom determines the maximum width

[16] We restricted the sample to less the 5 employees in order to allocate export orders over a larger number of firms.

of a rug a weaver can manufacture, and there was also uncertainty about the size of the rugs that Hamis would be able to secure. For reasons that will be come clear momentarily, we refer to these 303 firms as "Sample 1". The first two rows of columns 1-4 of Table 1 show the total number of firms by rug type and treatment status for Sample 1.[17]

We designed the following export-market access intervention. Hamis Carpets (with ATA's assistance) marketed rugs to overseas buyers and once export orders were secured we divided the order and allocated an *initial* amount to each of the producers in our treatment group. The treatment firms were visited by our survey team and a representative of Hamis carpets and offered the opportunity to the fill the 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 as a balance between a reasonable size order and the ability to have enough orders to treat the firms. We instructed Hamis to offer the market price based on the specifications of the rugs ordered by foreign clients (prices we analyze in detail below). If the firm accepted, Hamis delivered the input thread and the correctly sized reed and heddle to ensure all rug orders were consistent across producers. At the same time Hamis would discuss the technical aspects of the specific rug order and answer any questions the firm may have. Rather than a deadline, there was an implicit understanding that firms would deliver rugs to Hamis for payment on a weekly basis with payment upon delivery.

As further export orders were generated, Hamis continued to place them with the treatment firms. Just as in any arms-length transaction, after the initial order amounts were offered, Hamis was not bound to continue to make subsequent purchases from any particular treatment firm if the quality was below par or the previous rugs were not delivered on time. In other words, the experiment protocol simply forced Hamis to offer an *initial* order to the treatment firm. If the relationship did not work, Hamis was not forced to continue to offer orders to that firm. Hamis was not allowed to allocate any orders to control firms and we maintained a project coordinator and survey team in Fowa to ensure that the protocols were followed.[18] Thus, from the perspective of the treatment groups, the intervention provided them 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. As the anecdotes in Section 2.3 illustrate, Hamis' foreign buyers are demanding and would not accept subpar rugs. Second, and more importantly, for greater external validity, we wanted the experiment to mimic a normal buyer-seller relationship as closely as possible. Our intervention places initial orders with a random set of producers, but allows the intermediary to optimally allocate further orders *within the treatment group* based on firm quality, reliability and so forth. As such, subsequent orders are endogenous and we will be clear about the

[17]The randomization occurred at the rug-type and loom-size level and some strata were uneven leading to 149 treatment firms out of the sample of 303 firms.

[18]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.

interpretation of any findings that draw on variation in the size of post-treatment orders. 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.

An alternative experiment would be to provide our control firms with a similar quantity of rug orders but from domestic rather than foreign sources. We did not pursue this approach for reasons both theoretical and practical. From a theoretical point of view, trade models typically model exporting as a demand shock, sometimes with features distinct from domestic demand shocks. Increasing demand is also the primary motivation for many export facilitation policies (e.g., sending trade delegations, researching foreign markets, building export infrastructure such as ports or streamlining export regulations). Therefore, to assess the impacts of exporting, it is natural to include this central component. In terms of the practical limitations, if we were to provide equally-sized domestic orders it is unclear on what dimension they should be equal given the different profit margins and hours required per rug. Even then, given the limited local demand in Egypt during a period of political turmoil, it would not have been feasible to acquire anything like the $152,302 of orders that we have managed to generate on international markets.

3.2 Experiment Takeup

The third row of Table 1 shows the takeup status for Sample 1 (columns 1-4). In general, the takeup numbers were disappointing with 22 percent of the whole sample taking up the opportunity to export. For goublain and tups producers, the two rug types for which we obtained no orders, take-up rates are 10 and 19 percent, respectively. We expected low takeup values in these strata since these firms did not produce duble or kasaees rugs. Nevertheless, we attempted to treat these firms and very few were willing to switch rug types.[19] During the second survey round, we asked firms to list the reasons for refusing treatment. The goublain and tups panels in Table 2 confirm that the main reason for refusals among these firms was that the export rug order was not the suitable rug type.

In contrast we did have export orders for kasaees and duble rugs. Table 1 shows that among kasaees and duble rug producers take up was 26 and 38 percent, respectively, but the takeup rates were still relatively low. As previously mentioned, and shown in Figure 5, between December 2010 and May 2012, ATA and Hamis were unable to secure a large number of export orders even for duble rugs. As a result, we were unable to approach treatment firms in Sample 1 with the opportunity to produce 110m^2 in one go. Instead, we had to offer smaller orders of 20m^2 sequentially, or about two weeks of work. Because this initial order size was small, many firms were unwilling to work with us. The duble panel of Table 2 compiles survey responses from duble firms and shows that many were unwilling to jeopardize their existing relationships with intermediaries for a small amount of work.[20]

[19] The lack of switching across rug types suggests that the price was not high enough to compensate these firms for switching rug types.

[20] Several duble treatment firms reported that the export order was not the suitable rug type as they misreported duble as their primary rug type at baseline. Many kasaees producers were unwilling to accept the export order because the particular rug they were asked to produce was different from the kasaees rugs they usually make.

Since March 2013, Hamis' major buyers offered assurances that they would continue to place duble rug orders for the foreseeable future so it was possible to offer the opportunity to produce 110m^2 in one go. We therefore decided to draw a second sample of firms that just produced duble. Given that all our export orders were for duble rugs, this would increase the sample size of duble firms substantially. Additionally, given the larger order size, we expected higher takeup within duble producers. In February 2013 the survey team found an additional 140 firms that specialized in duble production and were not in the original listing exercise; we refer to these firms as "Sample 2". As with the initial sample, we stratified these firms on loom size and 35 firms out of the 140 were randomized into the treatment sample.[21]

Given the large export orders that Hamis had secured, we could now offer the full 110m^2 at once to the treatment firms in Sample 2 and we could also ensure that all the treatment firms in Sample 1 received their full 110m^2 allotment. As predicted, this large order has led to substantially higher takeup in Sample 2 firms. Column 5 of Table 1 reports treatment and takeup statistics for Sample 2: 32 out of 35 firms agreed to produce the export orders for Hamis.[22]

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 Carpets. As shown in Table 1, only 4 treatment firms, all in Sample 1, failed to secure additional orders from Hamis after the initial treatment. Two of the firms were unable to manufacture the export orders successfully while the remaining two firms had a falling out with the owner of Hamis Carpets. The fact that the overwhelming majority of firms were able to produce the orders successfully is itself interesting, and is likely related to the learning-by-exporting results we explore below. Going forward, we use the numbers in row 3 to define the sample of takeup firms.[23]

Given that we were only able to generate large and sustained export orders in one rug type, duble, and that very few firms outside of duble were willing to manufacture this rug type, we restrict our analysis of Sample 1 to the duble strata. (Sample 2 only contains duble producers.)[24] In terms of the analysis, Sample 2 has two advantages. First, as noted above, the second sample had much higher takeup rates since Hamis was able to offer large initial treatment orders all at once. This means that there was less potential selection among takeup firms which affects the interpretation of the treatment-on-the-treated specifications. Second, the treatment in Sample 2 is the treatment we intended when designing the experiment. Firms in Sample 2 were offered a large initial order followed up by continued orders if the initial order was filled satisfactorily; in contrast Sample 1 firms did not receive a reliable flow of orders until 1.5 years after the beginning of the study. This fact can be clearly seen from Figure 5 which superimposes the dates of the sur-

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

[22] 30 of the 35 Sample 2 treatment firms immediately took the offer upon treatment assignment in March 2013 and the remaining 2 firms began producing orders for Hamis in May 2014. This delay was due to capacity constraints on the side of Hamis Carpets.

[23] We count firms who were not successful in sustaining work as having taken up to be as conservative as possible when calculating treatment on treated effects.

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

vey rounds on Hamis' cumulative exports. For these reasons, we will present two sets of results, the first restricting the analysis to Sample 2 firms only (with 140 duble producers, 35 in treatment and 32 who took up), and the second pooling all the duble producers in Sample 1 and Sample 2 (the "Joint Sample" with 219 duble producers, 74 in treatment and 47 who took up). Since Sample 2 is our preferred sample, we focus our discussion on the results for Sample 2 and note any discrepancies with the Joint Sample when they arise.

3.3 Data

Data collection for each sample occurred in three phases: baseline, periodic follow-up surveys and endline. In both the baseline and endline we collected data on (a) firm production, (b) rug quality, and (c) household and demographic characteristics. In the follow-up surveys we only collected data on firm production and rug quality. The initial intention was that follow-ups surveys would be conducted quarterly but political turmoil in Egypt resulted in several unanticipated delays.[25] Table 3 shows the timeline of surveys for both samples.

The firm production module records production activity of firms for the month preceding the survey interview. We collect direct measures of profits, revenues, expenses, output quantity and prices, input quantity and prices, and total labor hours worked, as well as information on the rug designs (for example, the design pattern and the number of colors used).

Each survey round includes a module that recorded the quality of the rugs produced by treatment and control 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:[26] (1) Packedness; (2) Corners; (3) Waviness; (4) Weight; (5) Touch; (6) Warp Thread Tightness; (7) Firmness; (8) Design Accuracy; (9) Warp Thread Packedness; (10) Inputs; and (11) Loom.[27] Each measure is rated on a 1 to 5 scale, with higher numbers denoting higher quality.

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

[25] On three separate occasions we experienced delays of several months in getting permits to survey the firms in Fowa from Egypt's Central Agency for Public Mobilization and Statistics (CAPMAS) as this agency was not processing applications during periods of political turmoil.

[26] The first Sample 1 surveys recorded 6 quality metrics to which we subsequently added 5 additional metrics.

[27] Packedness measures the how well the rug holds together (poorly packed rugs can have holes); 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. 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.

We collected a third set of quality measures in June 2014 by setting up a "quality lab" in a rented workshop where firm owners were brought and asked to produce an identical domestic-specification rug using identical inputs and equipment. The rugs were then anonymized and scored along the quality dimensions listed above by both the master artisan and a Professor of Handicraft Science.

We administered a household module at baseline and endline. This module collects information on household income, literacy rates and so forth. Below we show that the samples are balanced across these dimensions as well.

Finally, we note that we hired an Egyptian survey company to conduct the baseline survey on Sample 1. The company also trained an enumerator who became responsible for regular short follow-up surveys on the firms. Unfortunately, we discovered that the enumerator they trained had made up much of the data for the first follow up round, Round 1, and so this data has been discarded. We immediately fired the enumerator and hired new employees who have been employed since January 2012 and conducted all subsequent surveys.

3.4 Summary Statistics

Table 4 shows baseline balance between the treatment and control groups for Sample 2 in the left panel and the Joint Sample in the right panel. The table reports regressions of each variable on a treatment dummy and strata fixed effects, and reports the constant (the mean of the control firms) and treatment coefficient (the difference between control and treatment means). Panel A shows summary statistics for the household characteristics of the firm owner. The first row reports that the mean age in treatment and control is around 50 years and row 2 indicates that, on average, firms have slightly more than 35 years of experience working in the rug industry. Roughly 60 percent of firm owners are illiterate. The average household size is 4.

Panel B reports statistics from the rug business. Monthly profits from the rug business averages LE734 in Sample 2 ($107.78 at the prevailing exchange rate) and LE548 in the Joint Sample ($92.10 at the prevailing exchange rate). Firms report 268 labor hours in the previous month, which amounts to around 22 days of work at 12 hours per day. As noted earlier, firm sizes are small because this was an explicit criterion in choosing our sample: the average firm has just over one employee. Total output per month is 43.5m^2 and only about 16 percent of firms have ever knowingly produced rugs for the export market. The final row of Panel B reports the average rug quality across the 11 dimensions and finds no statistical difference between treatment and control firms. In the Joint Sample, we do observe a statistical difference, but treatment firms report *lower* quality scores at baseline. The final row of the Table reports attrition across survey rounds. Attrition has been low with a non-response rate of approximately 4 percent per round (11 percent for the Joint Sample) which does not vary across treatment and control groups in either sample.

4 Causal Impacts of Export-Market Access on Profits

4.1 Empirical Specifications

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

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

where y_{it} is the profit measure, $Treatment_i$ is an indicator variable that takes the value 1 if the firm i is in the treatment group, τ_t are time period fixed effects, δ_s are strata fixed effects and y_{i0} is the value of the dependent variable at baseline. Since (1) controls for the baseline value of the dependent variable, we of course do not include observations from the baseline survey round in the regression. The coefficient on $Treatment_i$ provides the causal impact of being provided with the opportunity to export.[28]

Since, as discussed in Section 3, not all firms who were offered the opportunity to export took up that offer, (1) is the intent-to-treat (ITT) specification. We also present results from the treatment-on-the-treated specification (TOT) which scales the treatment effect to take account of the fact that not everyone was actually treated (under the assumption that the offer to export does not affect the outcome of interest for those who do not take up the treatment):

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

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 (and stays on) when a firm first produces carpets for the intermediary. Of course takeup is not random and may be correlated with unobservables, and so in order to obtain the TOT specification 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 treatment worked, in so far as treatment firms were more likely to knowingly have manufactured rugs for export markets. To show this we replace y_{it} with a dummy variable that takes the value 1 if a firm ever knowingly made rugs for export. These results are shown in Table 5. Being in treatment raises the probability of ever exporting by 68 percentage points from a baseline of 19 percent in Sample 2 and 54 percentage points from a baseline of 13 percent in the Joint Sample. We also report the TOT specification, where actually taking up our opportunity to export increases the probability of ever exporting even more dramatically, which is not surprising.[29]

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

[28]Alternatively we could use all survey rounds and include firm fixed effects in which case we would regress the dependent variable on $Treatment_i$ interacted with a dummy for post-baseline round since the opportunity to export was provided after the baseline. We chose our specification because if the dependent variable is measured with noise, the fixed effects estimator will perform poorly with a limited number of survey rounds.

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

firms. Their evidence suggests that attempts to construct firm profits from data on revenues and expenses may be noisy. 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 profits in a given month will not capture firm profits on material expenses. They advocate simply asking firms to directly report profits.

Following de Mel et al. (2009), we construct four measures of profits. The first measure is a direct profit measure from the firm's response to the question: "What was the total income from the rug business last month after paying all expenses (inputs, wages to weavers but excluding yourself). That is, what were your profits from this business last month?" The second measure constructs profits from two surveys questions that ask firms to report their total revenues and total costs from the previous month. The third measure constructs profits from the production modules that contain detailed information on prices and quantities of inputs and outputs. The idea behind this measure is that there may be less noise in constructing profits from its components–prices and quantities–than from recall information on total revenues and expenses; we refer to this measure as "constructed profits". This measure is also free of the concern that firms might use business expenses for household consumption (or use business revenues to pay for household expenses) that may be confounded in 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 measures alleviates concerns regarding the timing of when revenues are earned and costs are incurred and serves as a check against the three profit measures.

4.3 Profit Results

Table 6 shows the results of running the specifications above on the various profit metrics. The table is divided into four panels: the top two panels reports results using Sample 2 and the Joint Sample, respectively. As before, we discuss the results from our preferred sample, Sample 2, and note if there meaningful differences in the Joint Sample. The columns display different profit measures as outcome variables and for each we report the ITT and TOT specifications.

The first two columns (1A and 1B) of Panel A of Table 6 report the specifications using the (log) profit measure that we discussed in Section 4.2. The ITT coefficient is 0.25, implying that the export treatment increases monthly profits by approximately 25 percent. The TOT coefficient is, not surprisingly, larger at 30 percent and is also statistically significant. The ITT point estimate from the Joint Sample is of similar value and the TOT is higher at 42 percent.

Columns 2A and 2B of Panel A report specifications using a profit measure constructed from asking firms about total revenue and costs in the previous month. As noted earlier, it is possible that this measure may be noisier than the direct profits measure. However, we observe very similar point estimates: the ITT and TOT are 23 and 28 percent, respectively. The reason these point estimates are similar to column 1 may be because the firms in our sample typically do not store much inventory and hence the timing mismatch between revenues and expenses is not severe. Moreover, as discussed by de Mel et al. (2009), when asked about revenues and expenses, firms often fail to account for business expenses that are used for household consumption. This issue is also less likely to be a problem in our context since our sample comprises of manufacturing firms whose inputs are unlikely to be used for household consumption.

We report the results using constructed profits in the columns 3A and 3B. The ITT and TOT point estimates are again very similar to the previous columns: the opportunity to export raises profits by 24 percent. Finally, we examine the "hypothetical profit" measure in columns 4A and 4B. These estimates are higher than the previous numbers. The ITT point estimate is 36 percent. The first three measures are likely better measures of realized profits, so we put more faith in those estimates, but it is reassuring to see consistency across all four measures.

These regressions indicate that the export treatment causally increases profits by between 20-25 percent. Of course, profits may have risen if firms increased their employment hours. In Panels C and D of Table 6, we construct profits per hour by dividing each profit variable by reported total hours worked in the previous month. Using the direct profit per hour measure in column 1A and 1B, we find that the ITT estimate in Sample 2 (Panel C) is 17 percent. This estimate is lower than the corresponding estimates in columns 1A and 1B in the previous table which implies that treatment firms report working longer hours. The remaining columns also show slightly lower estimates as well. The differences between Panels A and C suggest that total hours increased by 6 to 8 percent, depending on the sample.[30] In the subsequent section, we examine this increase in hours in more detail. Nevertheless, the basic message remains the same: the opportunity to export raised profits per hour by 15-20 percent. This represents a sizable impact on profits through improved market access.

4.4 Discussion of Profit Results

The bulk of the papers studying the impact of trade on firm performance examine changes in productivity. However, we believe that the analysis of accounting profits is interesting in its own right. The World Trade Organization estimates that $48 billion is spent annually on Aid for Trade programs designed to improve the capacity of developing countries to integrate more effectively into the multilateral trade organization (WTO, 2013). Are these policies cost effective? Despite their pervasiveness, we know very little about the efficacy of these policy initiatives in improving firm performance, and if they are effective, whether or not these improvements occur through learning-by-exporting or another mechanism (Fernandes et al., 2011). Having shown the impacts on profits, the rest of this paper investigates the mechanisms through which export market access affects firm performance. In future work, we plan to draw on these results along with data on the costs incurred to successfully generate export orders in order to provide an explicit cost-benefit analysis of this particular program, as well as assessing the implications for welfare.

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

Before turning to mechanisms, 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 (see Table 9 of McKenzie and Woodruff, 2013).[31] One possible interpretation of the mixed results is that investments in management and production practices may only be effective in the absence of demand constraints. For example, the returns to business literacy may be low if there is insufficient demand. Our results suggest a potentially important role for relaxing demand constraints through expanding market access. Another popular intervention normally targeted at small firms is expanding access to credit. The literature on the impacts of credit on profits for small firms also finds mixed results. For instance, de Mel et al. (2008) find returns to capital of around 5 percent per month while Banerjee (2013) cites several credit interventions that produced no statistical increases in profits. As such, our evidence suggests that demand constraints may be a key factor limiting the growth of small firms.

5 Sources of Profit Changes

5.1 Prices, Output, Input Factors and Costs

This section uncovers the particular mechanisms driving the increase in profits we found. The literature highlights various channels through which increased market access can improve firm performance. To fix ideas consider the following profit function for a firm:

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

where p is the price a firm receives for one unit of rug. The quantity of rugs produced is x, w is the wage paid for each hour of labor l and F is a fixed cost of production.

Our survey data contain the components of the profit function necessary to decompose the profit increase shown in the previous section. We begin by evaluating the impact of the intervention on the output price, p. In our setting, 96 percent of Sample 2 firms (and 74 percent of firm in the Joint Sample) are provided raw material inputs from their intermediary and hence do not pay for these expenses (hence, we exclude input expenditures in (3)). Hamis Carpets follows this industry norm. For the small percentage of firms that do purchase inputs on their own account, we subtract the prices of the warp and weft thread inputs from p to make these prices comparable across all firms. We examine the impact on (log) prices using the ITT and TOT specifications in the first columns of Table 7. The ITT specification indicates 46 percent increase in prices with the opportunity to export while the TOT indicates a 56 percent increase. Thus, part of the profit increase from exporting is coming from significantly higher prices per m^2 of rug for export orders.

Columns 2A and 2B examine the impact of the opportunity to export on total output weaved

[31] Calderon et al. (2013) provide a 48-hour business skills course to a random set of female entrepreneurs in rural Mexico and find 23 percent increase in profits relative to the control firms. de Mel et al. (2014) find no evidence that business training alone improves profits among existing firms in Sri Lanka, but profits do rise about 17 percent among firms who randomly receive both business training and a cash grant, and on new firms who receive business training.

by the firm in the previous month (measured in m^2 and unadjusted for product specifications). We find a large *decline* in total (log) output. The ITT estimate is -22 percent while the TOT is -27 percent. The estimates suggest large output declines among treatment relative to control firms.[32]

We next document the impact of the intervention on firm scale, as captured by total employment hours l. Columns 3A and 3B of Table 7 report these scale results with (log) total hours worked by all employees in the firm in the previous month as the dependent variable. The ITT estimate indicates an increase of 8 percent and the TOT is 10 percent. Since most firm owners are the primary weavers, and helpers are often family members, we have limited data on the shadow wage w that may also be responding to the opportunity to export, although in the case of single-employee firms, that employee is the residual claimant to profits (and we showed profits per hour increased in Table 6).

Finally, we turn to fixed costs F in columns 4A and 4B. The main proxy we use to capture changes in fixed costs is the size of the warp thread ball, measured in (log) kilograms, that is placed on the loom at the beginning of a production run. A larger warp thread ball enables firms to amortize the costs of re-stringing the loom through longer production runs. The ITT estimate is 13 percent indicating that the treatment led to sizable increases in the initial size of the warp thread ball. Hence, the data suggest 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.

In Table 8, we examine 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. We also tracked their input quantity usage. The first two columns of Table 8 examine the impact of the intervention on reported weft and warp thread prices. Recall that the weft thread is used to create the pattern of the rug and the warp thread is the base thread that is not observable in the finished rug but is important for maintaining the rug structure. Reported weft thread prices increase 23 percent in Sample 2. In contrast, there is no evidence that warp thread prices are higher among treatment firms. These two findings are sensible given the production technology for rug making. The warp thread is a thin thread that provides the structure of the carpet. As such, there is little variation in the specifications of the warp thread across rugs compared to the weft thread, which can vary by material type (cotton, wool, polyester, silk or various blends), thickness and material grade (e.g., Egyptian wool or more expensive New Zealand wool). Columns 3-4 suggest that input quantities do not increase with the opportunity to exports, but given the decline in output noted above, it 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

[32]The sum of the point estimates on prices and output matches column 3 in the previous table for Sample 2. It does not exactly match column 3 for the Joint Sample because of missing price data for a handful of observations.

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 rug output with exporting that we find is not consistent with either of these frameworks. The results are also not consistent with exporting simply being a generic demand shock (which would 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 input, rug output may actually fall alongside increasing revenues and input usage. The rise in material input prices provide further evidence for such an explanation if high-quality rugs require more expensive high-quality inputs (Kugler and Verhoogen, 2012). In the next subsection we confirm this conjecture. That exports lead to improvements in quality for firms in developing countries has been suggested by Hallak and Sivadasan (2013), who argue that even small firms are able to break into export markets by offering high-quality products, and by Artopoulos et al. (2013) who provide detailed case studies of Argentinian exporters improving product quality.

5.3 Quality Upgrading and Quantity-Based Productivity

We first draw on the detailed quality metrics described in Section 3.3 to confirm that treatment firms are indeed manufacturing higher quality products. We have 11 different quality metrics that are ranked on a 1-5 basis with 5 being the best for that type of quality. Conversations with Hamis Carpets and buyers reveal that the first five quality measures—packedness, corners, waviness, weight and touch—are the most important determinants of rug quality.

Table 9 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 and both the strata and round fixed effects as well as a constant. 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 every quality metric except one, there is strong evidence that quality is higher among treatment firms relative to control. The ITT estimate on packedness is 1.38 and the TOT estimate is 1.68. We observe similar patterns for the other quality measures—shape of the corners, the waviness of the rug, the weight of the piece, the feel of the rug and so forth. 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 measures separately, in panel B of Table 9 we re-run the analysis by restricting the coefficients on the treatment dummy and controls to be identical across the various quality metrics (recall they were all run in a single stacked regression). Given the previous results, it is not surprising that we obtain positive and statistically significant ITT and TOT estimates when we do this. On average, quality (on a scale of 1 to 5) is 1.14 points higher among treatment firms. These are substantial increases in quality given a standard deviation of quality of 0.55 at baseline.

We also examine productivity, measured as output per labor hour. This productivity measure is based on the production technology described in Section 2.2. The production technology is Leontief in labor and materials. Labor is the primary input and materials are non-binding since the majority of firms are provided inputs by their dealers. We abstract from capital because there is very little variation in capital across firms: 92 percent of firms use one loom (and 98 percent in Sample 2) and no firm in our sample purchased (or rented) an additional loom since the beginning of the study.[33] That said, we also consider a second measure of productivity, output per unit input, that we construct using a production function that includes the number of looms.

We have two ways to measure productivity. The first is *output per hour* which we get from firms' responses to the question: "how long does it take you to make 1 meter squared?". [34] The second productivity measure relaxes the assumption that only labor is required for production and estimates TFP using a Cobb-Douglas production function with both labor and capital and accounts for simultaneity of input choices (see Appendix A for further details).

Panel A of Table 10 shows the ITT and TOT results for output per hour. The ITT estimate indicates that output per hour fell 24 percent among treatment firms relative to control firms with even larger TOT effects. The bottom panel presents the TFP measure and we find a similar 29 percent decline in the ITT specification.

These two findings on quality and productivity are consistent with learning-by-exporting generating the profit increases found in Section 5.1. If the ability to produce high-quality rugs rose through learning-by-exporting, and such rugs take longer to produce, we would expect to see rising profits and quality accompanied by declining productivity. However, we would see similar patterns if the opportunity to export simply raised the price of quality. We now turn to clarifying these two mechanisms.

5.4 Passive-Quality-Upgrading versus Learning-by-Exporting

There are two potential ways through which quality upgrading can occur, and the distinction is crucial for understanding how exporting improves firm performance. We call the first mechanism *passive-quality-upgrading.* In this story, firms always knew how to manufacture the high-quality rugs demanded by rich-country buyers. If foreign buyers pay higher prices, but particularly so

[33] Looms vary by size but since our measure of output is in terms of square meters there is no reason to expect output to vary depending on loom size, all else equal. In any case, we control for loom sizes through strata fixed effects in the analysis below.

[34] Another way to measure output per hour is to divide total output by total hours worked in the month. We believe that the latter response is measured with less noise and use this direct measure for our analysis. We find virtually identical results using a measure that divides total output by total hours (available on request).

for high-quality products, firms will upgrade quality as long as the returns offset any costs (e.g., more expensive inputs or more hours devoted to producing the more demanding specifications associated with high-quality rugs).[35] This is a movement *along* the production possibilities frontier. Such an outcome would naturally occur if demand for high-quality rugs is less elastic, or if there is less competition in this segment. While it is quite challenging to provide a direct mapping between markups and quality levels without imposing additional structure, we provide some suggestive evidence for this phenomenon by analyzing Hamis Carpets' (self-reported) cost structure for domestic and foreign orders. Hamis reports 9 percent markups on domestic orders and substantially higher markups of 33 percent on foreign orders with the full cost structure broken down in Appendix Table B.1. This provides some evidence that the higher prices we observe among treatment firms may come from this markup being shared between Hamis and the producer. In this passive-quality-upgrading story, the export opportunity raises the price of high-quality rugs and profit-maximizing firms respond by producing rugs to specifications associated with high-quality. What does *not* change in this passive story is technical efficiency.

We distinguish passive-quality-upgrading from *learning-by-exporting*, which we define as an export-induced change in technical efficiency. This is a shift *out* in the production possibilities frontier. If such changes in technical efficiency are biased towards high-quality production, quality upgrading can also occur through these learning processes. To be clear, the two channels are not mutually exclusive. The opportunity to export may both increase the price of quality and result in changes to the production function through knowledge flows or other learning-by-exporting mechanisms. 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 passive upgrading. Such learning processes are vital for generating dynamic gains from trade. In the next section, we turn to providing evidence for learning-by-exporting and demonstrate that our findings are inconsistent with a pure passive-quality-upgrading story.

6 Detecting Learning-by-Exporting

In order to better understand learning-by-exporting, and distinguish it from passive-quality-upgrading, we enrich the profit function by detailing production functions for output and quality:

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

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

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

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

where p is now a price function that is exogenous to the firm and depends on the quality of the rug q, with $b > 0$. Rug quantity and quality are determined by two production functions, both of which depend on a choice variable: the product specifications of the rug indexed by λ. High-λ

[35] For evidence that export quality covaries with destination income per capita, see Schott (2004), Hummels and Klenow (2005), Hallak (2006, 2010), Crozet et al. (2012), and Hallak and Sivadasan (2013).

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

The production function for output $x(\lambda, l)$ has two components. Labor inputs are mapped to output through $f(l)$ and output per unit of labor input is determined by the function $a(\lambda, \chi_a)$, an output productivity measure that is "unadjusted" for rug specifications. We chose this simple parametrization since, as discussed in Section 5.3, there is little variation in capital across firms and the production technology is Leontief in materials and labor, with labor the binding factor.

Output productivity $a(.)$ is necessarily decreasing in λ since rugs with more demanding specifications require more labor hours. The function $a(.)$ is also increasing in χ_a, an efficiency knowledge parameter. Collecting these two derivatives:

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

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

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

With this structure in hand, it is straightforward to clarify the distinction between passive-quality-upgrading and learning-by-exporting. We define passive-quality-upgrading as increases in b, the price of quality, that cause firms to choose higher λ (and hence higher quality). In contrast, we define learning-by-exporting as the process through which exporting raises χ_a and χ_q, the two knowledge parameters. This process can occur either as firms move into high-quality products with steep learning curves or through transfers of knowledge between foreign buyers and domestic sellers. We might expect transfers of knowledge about quality, χ_q, to be particularly relevant for firms in low-income countries that sell to buyers in high-income countries that have have more sophisticated tastes and demand higher-quality levels of production.[36] However, to our knowledge, systematic evidence showing that exporting causes improvements in knowledge about quality has not been documented in the literature.[37]

To see that this theoretical framework can generate output reductions alongside quality improvements consider the following simple functional forms for the production functions:

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

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

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

where λ is complementary with χ_a and χ_q in producing output and quality respectively (weakly so

[36] The case studies from Artopoulos et al. (2013) argue that this mechanism is important.

[37] There are, of course, papers that document quality upgrading by firms in developing countries as they export to richer countries (e.g., Verhoogen (2008) or Hallak and Sivadasan (2013)), but these findings are consistent with both passive-quality-upgrading and learning-by-exporting.

in the former case). Maximizing (4) using the functional forms in (10)-(12) yields the equilibrium product specifications and labor inputs, and hence equilibrium quality and productivity:

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

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

By taking derivatives of the equilibrium values with respect to either the price of quality (b) or the knowledge parameters (χ's) it is easy to see that both passive-quality-upgrading and learning-by-exporting can generate the results we found in the previous section. Increases in b or the quality knowledge parameter, χ_q, lead firms to raise product specifications and hence produce higher quality products that sell for higher prices. However, output and output productivity may fall alongside rising labor demand as high-λ rugs require more labor inputs.

Empirically detecting learning-by-exporting is challenging for two reasons. First, firms with high knowledge 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 understand and can 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 changes in technical efficiency through residual-based total factor productivity. If prices are higher in export markets, productivity measures that do not adjust for prices (which is rarely the case) may suggest learning-by-exporting when there is passive-quality-upgrading or simply a higher markup.[38] In the few cases where price adjustments are made, measuring quantity-based productivity requires comparing products with identical specifications. This is typically achieved by either relying on product dummies (e.g., de Loecker et al., 2014), with the extent of disaggregation determined by administrative classifications, or focusing on homogenous goods like concrete and block ice (e.g, Foster et al., 2008). In contrast, we exploit our rich panel data that contain both product specifications and quality metrics.

We test four implications of our simple model to distinguish learning-by-exporting from a passive-quality-upgrading story:

1. In Step 1, we show that although output productivity falls with the opportunity to export, productivity *conditional* on rug specifications *rises*. We also show that quality levels rise conditional on rug specifications. Under a pure passive upgrading story, output productivity and quality levels should be *unchanged* once we condition on λ.

2. In Step 2, we demonstrate that when asked to produce *identical* domestic rugs, treatment

[38] See de Loecker (2011) for an extensive discussion of this point. Marin and Voigtlander (2013) use the methodology developed by de Loecker et al. (2014) to purge productivity measures of prices and find export-induced efficiency gains, but their results rely on propensity score matching techniques, which requires fully specifying the selection model.

firms produce higher quality products and do not take longer to do so. Under passive-quality-upgrading, quality and productivity should not differ across treatment and control firms when making identical domestic rugs.

3. In Step 3, we use time-series data to establish that quality and output productivity evolve over time as cumulative export production increases, consistent with a learning process. In contrast, a passive upgrading story predicts a discontinuous jump upon exporting as firms immediately move to new quality and output productivity levels.

4. In Step 4, we show that quality is rising, in part, through the transfer of knowledge between foreign buyers, intermediaries and producers. If quality upgrading was passive, quality should be determined by prices alone. This knowledge may be about a technique (for example, how to make flatter carpets) or about what high quality entails (for example, that delivered sizes cannot deviate from the precise size specified by the buyer). To provide evidence of actual knowledge transfer, we draw on records of information flows including the precise topics discussed between the buyers and Hamis Carpets, and between Hamis Carpets and the firms. This evidence also allows us to distinguish between two forms of learning-by-exporting, transfers of knowledge (learning-from-others) and knowledge accumulation through production (learning-by-doing).[39]

6.1 Step 1: Conditioning on Rug Specifications

In a passive-quality-upgrading story, changes to quality levels and output productivity should occur only through changes in rug specifications. After conditioning on these specifications, quality and productivity should remain unchanged: $\frac{da}{db}|_{\lambda} = 0$. 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 (conditional on rug specifications) to rise due to an increase in χ_a, $\frac{da}{d\chi_a}|_{\lambda} > 0$. We also explore a related prediction for quality: in a passive-quality-upgrading story, quality should not change conditional on specifications, $\frac{dq}{db}|_{\lambda} = 0$, while under learning-by-exporting quality should rise with χ_q even conditioning on rug specifications, $\frac{dq}{d\chi_q}|_{\lambda} > 0$.

In order to test these distinguishing predictions, we repeat the quality and productivity regressions above but now control for the specifications of the rug being manufactured at the time of the survey visit.[40] Although we are not able to control perfectly for the myriad of possible rug specifications, we include five sets of controls that capture the key rug specifications: These five controls are (1) the type of rug being produced, (2) how difficult the rug was rated on a 1-5 scale by the master artisan, (3) the amount of thread used per m^2 of the rug, (4) the number of colors used in the rug, and (5) which segment of the market the rug is aimed at as reported by the master artisan

[39] For example, foreign buyers may demand rug specifications that firms have not produced prior to exporting and so firms have not already acquired the knowledge required to make them quickly. In this sense, learning through production is ultimately triggered by export orders and hence we consider this as a learning-by-exporting channel.

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

(normal, mid, or high). For example, rugs that are destined for lower tier stores would be labeled as "low". Including these controls is possible because there is overlap in rug specifications across firms selling to domestic and foreign markets. This overlap can be clearly seen in Figure 6 which plots the distribution of each of the five specifications both for firms that are producing rugs for export (i.e. $Takeup_{it} = 1$) and for those that are not. Note that if our characteristic controls are very crude, that will tend to bias our findings towards the unconditional results we found in Tables 9 and 10. Hence, the prediction that output productivity should rise conditional on specifications is particularly informative since output productivity fell in the absence of controls.

We explore the stacked quality measures in Panel A of Table 11. Focusing on Sample 2 in columns 1-2, we see that the ITT and TOT estimates are positive and statistically significant. The positive coefficient on the treatment dummy is evidence of learning-by-exporting. Again, if there was only passive-quality-upgrading, we would not observe improvements in quality *conditional* on product specifications. Moreover, the product specifications have the intuitive signs. More difficult rugs are associated with higher quality while those destined to lower segments of the market are associated with lower quality. We present the results for each quality metric in Appendix Table B.2 (left panel) and the conclusions are unchanged.

The bottom panels of Table 11 reports the results for productivity. Notice that conditioning on rug specifications *flips* the sign on the treatment dummy from negative to positive. That is, conditional on the five key rug specifications, the opportunity to export is now associated with significantly higher output productivity. Recall that Panel A of Table 10 reports an ITT estimate of *negative* 24 percent, and this changes to *positive* 28 percent when we condition on rug specifications in Panel B of Table 11. The TOT estimates exhibit a similar reversal, as do all the estimates using TFP in Panel C. Under pure passive-quality-upgrading, there should be no changes in output productivity conditional on product specifications. The data strongly suggest otherwise. As before, the rug characteristic controls have the intuitive signs: rugs with more colors and those that require more thread per m^2 take longer to weave (lower output per labor input). Relative to products at the high-end market segment, lower segment rugs take less time to weave. The adjusted R-squared nearly triples (from 0.18 to 0.53 in the ITT estimates) suggesting that the rug specifications have substantial explanatory power.

A reasonable concern with this exercise is that, although the treatment variable is exogenous by design, the characteristic controls are not necessarily exogenous yet are included as independent variables. An alternative approach is to adjust the productivity and quality measures for specifications, and regress the specification-adjusted measures on treatment. To perform this adjustment, we first regress productivity (or quality) on the five rug specifications at baseline, before any experimental intervention, and use the resulting coefficients to construct adjusted productivity (actual minus predicted productivity) for each round.[41] Of course, if higher-ability firms selected into higher-specification rugs at baseline, the coefficients on rug specifications will be bi-

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

ased. Specifically, the productivity penalty for making a high-specification rug is likely to be larger than the coefficients imply due to selection. If our experiment induced lower-ability firms to make high-specification rugs, the ITT comparing specification-adjusted productivity between treatment and control is likely to be biased downward (because the lower-ability firms in the treatment group "appear" less productive since we used potentially biased coefficients for the adjustment). Table 12 shows the results of regressing the specification-adjusted quality and productivity measures on treatment (or on takeup instrumented with treatment).[42] Reassuringly, the results are both qualitatively and quantitatively similar.

6.2 Step 2: Production of Identical Domestic Rugs

The second test exploits our experimental setting and compares quality and productivity across firms producing identical domestic rugs rather than relying on specification controls. If quality upgrading was passive, 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 workshop in June 2014 and asked them to produce an identical domestic-specification rug using the same loom. We chose domestic rug specifications that mimicked rugs sold at mid-tier domestic retail outlets.[43] We provided both inputs and the loom and the firm owner was paid a generous wage of LE40 to compensate for having to make the rug at an external location.

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 dimensions discussed above.[44] 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 performed an additional quality survey that serves as a further audit of the master artisan's quality scoring that we rely on in previous sections. We sent each of the rugs to be graded by a Professor of Handicraft Science at Domietta University located 2 hours east of Fowa. Having an external assessor from outside of Fowa provides a cross check on the accuracy of the master artisan's scoring.

In Panel A of Table 13, we report results separately for each quality metric. As before, we account for correlations across quality metrics by clustering standard errors by firm and reporting the treatment coefficients interacted with each quality metric; these coefficients are identical to running the regression separately for each quality metric. Quality is higher among the treatment firms and statistically different from control firms for every quality dimension. Reassuringly, the same is true using the professor's assessments: treatment firms consistently score higher along every quality dimension.

Panel B of Table 13 reports the results constraining the ITT and TOT estimates to be the same across quality metrics. Given the results from Panel A, it is not surprising that the coefficients are

[42]The right panel of Appendix Table B.2 reports the estimates separately for each adjusted quality metric.

[43]The rug design is popular in the domestic market an measures 140cm by 70cm with a desired weight of 1750 grams. The master artisan assigned a difficulty rating of 3 for this rug (below the 4.28 average rating of export orders).

[44]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.

statistically significant. The point estimate from column 1A of Panel B is 0.86 (and statistically significant). Given that the standard deviation of quality metrics from the master artisan's rating is 0.84, the point estimate suggests that quality levels of treatment firms are slightly more than a full standard deviation above control firms.

Panel C of Table 13 reports how accurately rugs match the length, width and weight that we requested. We define these variables as the minus of the absolute deviation from the target value, so higher values indicate more accuracy. Treatment firms produce rugs with more accurate rug lengths. We do not observe statistical differences in the width of the rugs, but this is expected since the loom size determines the width and the loom was the same for all firms in the lab. The third row reports weight accuracy. Once again, treatment firms produced rugs that were closer to the desired 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 accurately reflects firm productivity. The 4th row of Panel C reports the time taken, measured in minutes. On average, firms took 4 hours to produce the rug and there is *no* difference in time taken across treatment and control firms. That is, despite manufacturing rugs with higher quality metrics, treatment firms did not spend more time weaving: quality per unit time is higher.

Under a pure passive-quality-upgrading story, we would not expect these patterns of results. Instead, there should be no quality differences between exporting and non-exporting firms when asked to produce identical domestic rugs using the same inputs and capital and at the same scale. If anything we might expect non-exporting firms to produce these rugs quicker or at higher quality since they are more used to manufacturing domestic designs and specifications. In contrast, we find strong evidence of higher quality levels among treatment firms that persist even when manufacturing rugs for the domestic market. Moreover, treatment firms exhibit no loss in productivity as they do not take longer to produce these rugs despite them being higher quality.

6.3 Step 3: Learning Curves

The third test relates to the time paths of quality upgrading. In a passive upgrading story, quality upgrading is instantaneous, while learning processes typically take time. Accordingly, we explicitly write the two knowledge parameters in period t as:

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

where $\mathbf{1}\left[\text{export}_{it}\right]$ is a dummy that takes the value of one if the rug output is for export. In this formulation, the knowledge parameters change with the opportunity to export through the cumulative production of export rugs. This captures the idea that knowledge improves with repeated interactions with buyers or potentially because learning curves are steeper among export rugs that are less familiar to the firms. Under a pure passive-quality upgrading story, $\chi_{k,t>0} = \chi_{k,0}$ for $k = q, a$. We distinguish between the two mechanisms by examining how output productivity and quality evolve over time. If results are driven by passive-quality-upgrading, since firms previously knew how to make high-quality rugs, quality should *immediately* jump and (unadjusted) productivity should fall with the first export order and remain at those levels subsequently. If there is learning-by-exporting, productivity and quality should rise with cumulative exports.

To investigate potential learning curves in a non-parametric manner, we carry out a two-stage procedure. In the first stage, we regress our variable of interest on firm fixed effects as well as round fixed effects.[45] In the second stage, we plot a kernel-weighted local polynomial regression of the residuals against cumulative export production. Since cumulative export production is only available for takeup firms, we just include these firms in the second stage (although all firms are included in the first stage when we de-mean by survey round). As previously, we focus on Sample 2 firms.[46] We also show similar plots that either include product specifications in the conditioning set in the first stage, or that use the specification-adjusted measures described in Section 6.1 as the dependent variable in the first stage. Finally, Appendix Figure B.1 presents similar plots using the partially linear panel data estimator proposed by Baltagi and Li (2002).

Figure 7 shows these residual plots for the two productivity measures as well as the stacked quality measure. The upper left graph reports the residual output per hour measures (not controlling for product specifications); the figure indicates a decline in output productivity until about 400 m^2 after which output-per-hour starts to rise. We draw similar conclusions from the TFP measure (middle-left figure). The two panels to the right include product-specification controls or use specification-adjusted productivity. In these cases, we do not observe a decline in output productivity consistent with Table 11, suggesting that the dips observed in the left panels are driven by moving to the more difficult product specifications demanded by foreign buyers.[47] In fact, both conditional productivity and specification-adjusted productivity rises with total export production, inconsistent with a pure passive-quality-upgrading story.

The bottom row of Figure 7 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 three months. On the right two panels, we observe a similar path when including product-specification controls or using specification-adjusted quality. This implies that the quality curve is not simply driven by changes in product specifications. Relative to productivity, the figures suggest much faster learning about quality χ_q than about efficiency χ_a.[48]

The quality measures reported in Figure 7 were recorded by the master artisan at the time of

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

[46] For Sample 1 firms, given the large gaps and small orders in the first year of the project, cumulative production is a far more noisy measure of potential knowledge transfers.

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

[48] The finding that learning about quality occurs quickly is consistent with recent estimates of learning documented in the literature. In a randomized control study of management practices, Bloom et al. (2013) find a very large separation in quality defects after just 10 weeks between Indian textile firms that implement modern management practices and those that do not (see Figure IV of Bloom et al., 2013). 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.

each survey. For the subset of firms that produced orders for the intermediary, we have additional, though limited, quality metrics that were recorded for each order delivered by the firm (often weekly). This gives us order-specific quality measures for each firm, and hence many more observations per firm than we have for the master artisan quality measures which we collected at every survey round. We can produce similar plots for the 6 high-frequency quality measures recorded in this manner. Figure 8 shows local polynomial regressions of the residuals for these 6 measures (after regressing each measure on firm fixed effects) with Appendix Figure B.2 presenting similar plots using the partially linear panel data estimator. For 4 of the 6 metrics (packedness, firmness, size accuracy, design accuracy) we observe similarly quick learning curves as in the master artisan data. For the fifth metric, the readiness of the rug for delivery (5 being the most ready and 1 the least), we observe learning but at a slower rate (and with large standard errors). Finally, the learning curve for weight accuracy (defined as the negative of the absolute value of the difference between the actual weight and the weight specified by the buyer) shows more limited evidence of learning.

To conclude, inconsistent with passive-quality-upgrading, we find evidence of learning over time.

6.4 Step 4: Knowledge Transfers

The results in Steps 1-3 indicate learning-by-exporting is present in our data. In this final step, we provide a further test that also allows us to distinguish between the types of learning-by-exporting. Specifically, the results so far are consistent with both knowledge transfers between buyers, intermediaries and producers and a story where learning curves are particularly steep for high-quality items demanded by foreigners. It is of course likely that both channels are occurring, and in this subsection we provide evidence that some of the quality changes we observe can be explained in part by knowledge flows.

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 Carpets, as well as between Hamis and the producing firms. The data on flows between buyers and Hamis Carpets are more suggestive in nature. Hamis Carpets has shared email communications with its foreign buyers, and earlier we used excerpts to document the challenges Hamis experienced in securing overseas orders. We provide additional excerpts documenting information flows about specific aspects of rug quality between overseas buyers and Hamis Carpets. In one correspondence, a foreign buyer complained that the packing of the rug was too tight and strong which results in "wrinkly" rugs:

> ... 1) the colours in the borders were mixed and not like the order as you see at the picture and copy of the order 1590 in the attachment! 2) the carpet is completely weave and wrinkly. We told you time before, your packing is to tied and strong, so we have always this problems and complains. 3) there are stripes every approx. 50 cm all over the rug?

On a separate occasion, the same buyer also noted that the edges of some carpets had frayed:

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

These conversations suggest that buyers are passing along both information on how to manufacture high-quality rugs (i.e., packing that is not too tight or strong) as well as information on what exactly high-quality products entail. For example, while it may be clear that a delivered rug should be completely flat, the foreign buyer also stresses that long-term durability is also an important dimension of quality.

We have more detailed data on information flows between the intermediary and the firms. Hamis Carpets documented the visits made to each of the treatment firms as well as the subject discussed during that visit. In particular, Hamis Carpets provided data on the total number of conversations, their average length and the topics of discussion over the project period.[49] The topics are categorized according to 10 quality metrics (the intermediary did not discuss input quality since it provided the inputs). These visits entailed discussions about production techniques to achieve higher quality along these dimensions as well as passing on feedback from the foreign buyers. Appendix Table B.3 summarizes these data. All of the firms that took up the opportunity to export were visited at least seven times, with the intermediary visiting the average firm about 10 times. Each visit lasted slightly less than a half an hour on average. Issues related to design accuracy, the weight of the rug and the tightness of the warp thread were discussed on at least half of the occasions. Other dimensions that were discussed regularly were the corners of the rugs and the firmness of the rugs.

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 we recorded in the most recent survey. 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 in the last survey round:

$$Quality_{id} = \alpha + \beta_1 Takeup_i \times \mathbf{1}[Talked\_About\_Dimension]_{id} + \beta_2 Quality_{id0} + \delta_i + \delta_d + \varepsilon_{id}. \quad (16)$$

We include firm fixed effects δ_i so that we explicitly compare across quality dimensions d within a firm. We also include quality metric fixed effects δ_d to control for different means across dimensions.[50] This regression asks if the changes in the dimensions of quality that were discussed were relatively larger than the changes in the dimensions that were not discussed. A significant β_1 coefficient is suggestive of a knowledge transfer story rather than a learning-by-doing story

[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).

where quality improvements come about through cumulative production alone. A significant β_1 is also inconsistent with a passive-quality-upgrading story, where quality would be independent of knowledge flows.

Table 14 reports the results. Using either sample, we find a positive and statistically significant association between the movements in quality, relative to baseline, and the indicator variable for whether the intermediary discussed that quality dimension with the firm. Columns 3-4 re-run (16) accounting for rug specifications and the results are unchanged. In Columns 5-6, we use the specification-adjusted quality metrics as the dependent variable and results are also unchanged. The positive correlation is strongly suggestive that knowledge is flowing from the intermediary to the firm. It is hard to completely dismiss alternative explanations that involve slow learning about what firms can get away with or the rug preferences of foreigners that are communicated through these discussions. However, takeup firms report that 91.7 percent of these discussions involved the intermediary providing "information on techniques to improve quality" (as opposed to only pointing out flaws). 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.[51] Combined with the conversations discussed earlier, we interpret these results as evidence in favor of learning-by-exporting occurring, at least in part, through direct transfers of knowledge.

We provide one additional piece of evidence that suggests that our results are not driven entirely by a pure learning-by-doing story. Under learning-by-doing we would expect firms who were already producing high-quality rugs at baseline to see smaller treatment effects since they had less to learn. This prediction is not borne out by the data. When we regress the stack of quality metrics on a treatment dummy, baseline quality and an interaction of the two, the coefficient on the interaction is insignificant.[52]

This section presented four pieces of evidence in support for a learning-by-exporting mechanism: evidence of quality and productivity upgrading after conditioning on rug specifications or when making identical domestic rugs, evidence of gradual improvements in quality and productivity over time, and evidence that quality responds to knowledge transfers.

7 Conclusion

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

[51] 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.

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

We find that operating profits for treatment firms increase 15-25 percent relative to control. This finding stands in contrast to many RCTs designed to alleviate supply-side constraints for small firms that have shown no such positive profit impacts. This suggests that demand-side constraints may be a critical barrier to firm growth in developing countries and can be potentially alleviated through market access initiatives. The rise in profits is driven by substantial quality upgrading accompanied by declines in quantity-based productivity, indicating that foreign buyers demand higher quality products that take longer to manufacture.

The quality upgrading we observe can occur through two distinct mechanisms. Under passive-quality-upgrading, firms always knew how to produce high-quality products but optimally chose not to do so because domestic buyers did not demand high-quality (or were unwilling to pay for it). In this mechanism, exporting leads to no changes in the technical efficiency of firms. In contrast, under learning-by-exporting, exporting induces improvements in technical efficiency biased towards the manufacture of high-quality rugs.

We provide evidence that passive-quality-upgrading alone cannot account for our findings. 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.

Taken together, the evidence indicates that learning-by-exporting is present in our data. 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 a particular 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 understanding the impacts of trade in developing countries.

References

Ackerberg, D., K. Caves, and G. Frazer (2006). Structural identification of production functions. Technical report, UCLA.

Ahn, J., A. K. Khandelwal, and S.-J. Wei (2011). The role of intermediaries in facilitating trade. *Journal of International Economics 84*(1), 73–85.

Artopoulos, A., D. Friel, and J. Hallak (2013). Export emergence of differentiated goods from developing countries: Export pioneers and business practices in Argentina. *Journal of Development Economics 105*(C), 19–35.

Aw, B. Y., M. J. Roberts, and D. Y. Xu (2011). R&D Investment, Exporting, and Productivity Dynamics. *American Economic Review 101*(4), 1312–44.

Baltagi, B. H. and D. Li (2002). Series estimation of partially linear panel data models with fixed effects. *Annals of Economics and Finance 3*(1), 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*(1), 487–519.

Bernard, A. B. and B. Jensen (1999). Exceptional exporter performance: cause, effect, or both? *Journal of International Economics 47*(1), 1–25.

Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does Management Matter? Evidence from India. *The Quarterly Journal of Economics 128*(1), 1–51.

Brambilla, I., D. Lederman, and G. Porto (2012). Exports, Export Destinations, and Skills. *American Economic Review 102*(7), 3406–38.

Calderon, G., J. M. Cunha, and G. D. Giorgi (2013). Business literacy and development: Evidence from a randomized controlled trial in rural mexico. Technical report.

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*(3), 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*(2), 609–644.

de Loecker, J. (2007). Do exports generate higher productivity? Evidence from Slovenia. *Journal of International Economics 73*(1), 69–98.

de Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica 79*(5), 1407–1451.

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*.

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*(4), 1329–1372.

de Mel, S., D. McKenzie, and C. Woodruff (2009). Measuring microenterprise profits: Must we ask how the sausage is made? *Journal of Development Economics 88*(1), 19–31.

de Mel, S., D. McKenzie, and C. Woodruff (2010). Wage subsidies for microenterprises. *American Economic Review 100*(2), 614–18.

de Mel, S., D. McKenzie, and C. Woodruff (2014). Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics 106*, 199–210.

Eaton, J., M. Eslava, D. Jinkins, C. Krizan, and J. Tybout (2013). A search and learning model of export dynamics. Technical report, Penn State.

Feenstra, R. C. and G. H. Hanson (2004). Intermediaries in Entrepot Trade: Hong Kong Re-Exports of Chinese Goods. *Journal of Economics & Management Strategy, Wiley Blackwell 13*(1), 3–35.

Feenstra, R. C. and J. Romalis (2014). International prices and endogenous quality. *The Quarterly Journal of Economics 129*(2), 477–527.

Fernandes, A. M., A. Mattoo, J. Gourdon, and O. Cadot (2011). *Impact Evaluation of Trade Interventions: Paving the Way*. The World Bank.

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425.

Grossman, G. M. and E. Helpman (1993). *Innovation and Growth in the Global Economy*, Volume 1 of *MIT Press Books*. The MIT Press.

Hallak, J. and P. K. Schott (2011). Estimating Cross-Country Differences in Product Quality. *The Quarterly Journal of Economics 126*(1), 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*(1), 53 – 67.

Hallak, J. C. (2006). Product quality and the direction of trade. *Journal of International Economics 68*(1), 238–265.

Hallak, J. C. (2010). A Product-Quality View of the Linder Hypothesis. *The Review of Economics and Statistics 92*(3), 453–466.

Harrison, A. and A. Rodriguez-Clare (2010). *Trade, Foreign Investment, and Industrial Policy for Developing Countries*, Volume 5 of *Handbook of Development Economics*, Chapter 0, pp. 4039–4214. Elsevier.

Hummels, D. and P. J. Klenow (2005). The variety and quality of a nation's exports. *American Economic Review 95*(3), 704–723.

Keller, W. (2004). International technology diffusion. *Journal of Economic Literature 42*(3), 752–782.

Khandelwal, A. (2010). The long and short (of) quality ladders. *The Review of Economic Studies 77*(4), 1450–1476.

Krugman, P. R. (1979). Increasing returns, monopolistic competition, and international trade. *Journal of International Economics 9*(4), 469–479.

Kugler, M. and E. Verhoogen (2012). Prices, Plant Size, and Product Quality. *Review of Economic Studies 79*(1), 307–339.

Levinsohn, J. and A. Petrin (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *Review of Economic Studies 70*(2), 317–341.

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*(4), 643 – 681.

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*(2), 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*(6), 1695–1725.

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

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*(4), 822–842.

Rauch, J. E. (1999). Networks versus markets in international trade. *Journal of International Economics 48*(1), 7 – 35.

Rauch, J. E. and V. Trindade (2002). Ethnic chinese networks in international trade. *Review of Economics and Statistics 84*(1), 116–130.

Schott, P. K. (2004). Across-product versus within-product specialization in international trade. *The Quarterly Journal of Economics 119*(2), 647–678.

UNCTAD (2010). Creative economy report 2010. Technical report.

Van Biesebroeck, J. (2005). Exporting raises productivity in sub-Saharan African manufacturing firms. *Journal of International Economics 67*(2), 373–391.

Verhoogen, E. A. (2008). Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector. *The Quarterly Journal of Economics 123*(2), 489–530.

Wagner, J. (2007). Exports and productivity: A survey of the evidence from firm-level data. *World Economy 30*(1), 60–82.

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

WTO (2013). Aid for trade at a glance: 2013. Technical report, World Trade Organization.

Appendix A Measuring Productivity

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

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

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

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

where x_{it} is the output (in m^2) of firm i at time t, l_{it} is total hours, k_{it} is the number of active looms, and a_{it} is firm productivity. We emphasize that there is virtually no variation in the number of looms across firms (92 percent of firms report having more than one loom), but we nevertheless allow the production function to depend on capital. The vector of controls, $\mathbf{Z}_{ft}$, include rug specifications, round and strata fixed effects. The ν_{ft} is an i.i.d. error term capturing unanticipated shocks and measurement error.

Although having quantity information and rug specifications deals with measurement concerns, there is still a potential simultaneity bias in estimating (A.1) since productivity is observed by the firm, but not us. The standard approach in the literature addresses simultaneity using the control function approach developed by Olley and Pakes (1996) and Levinsohn and Petrin (2003). We assume that capital is a dynamic variable subject to adjustment costs and labor is flexible. Material demand is given by $m_{it} = f_t(a_{it}, k_{it})$ and can be inverted as $a_{it} = f_t^{-1}(m_{it}, k_{it})$. We follow the literature and assume that productivity follows a first-order Markov process. We leverage the experimental setting by estimating the production function using only the control firms. This allows us to avoid parametric (or semi-parametric) assumptions on the productivity process of treatment firms, which we argue evolves with treatment over time in potentially non-linear ways.[53] We estimate the production function using the one-step approach proposed by Wooldridge (2009), with l_{it-1} as the instrument for l_{it}, and cluster standard errors by firm.[54] We obtain $\alpha_l = 0.77$ (s.e. of

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

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

.37) and $\alpha_k = 0.23$ (s.e. of 0.97). Given that the coefficients sum to 1, we cannot reject that there are constant returns to scale.[55]

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

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

Table 1: Firm Sample and Takeup Statistics

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

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

Table 2: Reasons for Refusing Treatment, Sample 1

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

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

Table 3: Survey Timeline

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

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

Table 4: Baseline Balance

	Sample 2			Joint Sample		
	Constant	Treatment	N	Constant	Treatment	N
Panel A: Household Characteristics						
Age	50.0	2.8	139	50.7	0.9	218
	(1.1)	(2.2)		(0.9)	(1.6)	
Experience	36.3	1.9	136	37.6	0.2	213
	(1.2)	(2.5)		(1.0)	(1.7)	
Illiterate?	0.57	0.07	135	0.59	0.10	214
	(0.05)	(0.10)		(0.04)	(0.07)	
Household Size	4.0	0.1	140	4.2	0.0	219
	(0.2)	(0.3)		(0.1)	(0.2)	
Panel B: Firm Characteristics						
Profits from rug business	734	-14.1	139	548	-30.1	218
	(34.7)	(72.9)		(73.7)	(108.0)	
Hours worked last month	268	1.3	139	247	-1.7	218
	(5.9)	(10.8)		(7.0)	(11.7)	
Number of employees	1.00	-	139	1.09	0.0	218
	-	-		(0.0)	(0.1)	
Total product last month (m^2)	43.5	0.33	139	48.9	3.3	218
	(2.7)	(5.81)		(4.8)	(10.0)	
Ever exported?	0.16	0.03	140	0.11	0.02	219
	(0.04)	(0.08)		(0.03)	(0.05)	
Average Quality	2.57	-0.09	140	2.68	-0.13 ***	218
	(0.03)	(0.06)		(0.03)	(0.05)	
Joint F-test		0.86			0.85	
Attrition	0.04	-0.02	420	0.11	0.00	815
	(0.01)	(0.02)		(0.01)	(0.02)	

Notes: Table presents baseline balance for the Joint Sample (left panel) and 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. The final row reports attrition statistics across survey rounds. Significance * .10; ** .05; *** .01.

Table 5: Impact of Intervention on Firms Knowingly Exporting

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

Notes: Table regresses an indicator 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. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 6: Impact of Exporting on Firm Profits

Panel A: Profits for Sample 2

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.25 ***		0.23 ***		0.24 ***		0.36 ***	
	(.06)		(.05)		(.05)		(.10)	
Takeup		0.30 ***		0.28 ***		0.29 ***		0.44 ***
		(.07)		(.06)		(.07)		(.12)
R-squared	0.30	0.31	0.28	0.29	0.29	0.29	0.34	0.35
Observations	375	375	375	375	375	375	373	373

Panel B: Profits for Joint Sample

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.26 ***		0.21 ***		0.19 ***		0.37 ***	
	(.05)		(.06)		(.06)		(.11)	
Takeup		0.42 ***		0.37 ***		0.33 ***		0.68 ***
		(.08)		(.10)		(.09)		(.19)
R-squared	0.23	0.24	0.18	0.19	0.20	0.23	0.21	0.21
Observations	573	573	644	644	656	656	687	687

Panel C: Profit per Hour Hour for Sample 2

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

Panel D: Profit per Hour for Joint Sample

	Direct Profits		(Reported Revenues - Reported Costs)		(Constructed Revenues - Constructed Costs)		Hypothetical Profits	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.19 ***		0.16 ***		0.15 ***		0.25 ***	
	(.05)		(.05)		(.05)		(.06)	
Takeup		0.31 ***		0.28 ***		0.26 ***		0.46 ***
		(.08)		(.09)		(.09)		(.12)
R-squared	0.17	0.17	0.15	0.15	0.16	0.17	0.23	0.21
Observations	573	573	644	644	655	655	687	687

Notes: Table reports treatment effects on different profit measures, all measured in logs. See text for details regarding each measure. Panel A runs ITT and TOT regressions on Sample 2. Panel B reports the analysis using the Joint Sample. Panels C and D report the analogous regressions using profits per hour as the dependent variable. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 7: Sources of Changes to Firm Profits

	Output Prices		Output (m^2)		Hours Worked		Warp Thread Ball (kg)	
Panel A: Sample 2								
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.46 ***		-0.22 **		0.08 ***		0.13 **	
	(.10)		(.09)		(.02)		(.05)	
Takeup		0.56 ***		-0.27 ***		0.10 ***		0.15 **
		(.12)		(.10)		(.03)		(.06)
R-squared	0.27	0.28	0.19	0.19	0.12	0.12	0.20	0.20
Observations	376	376	375	375	375	375	377	377
Panel B: Joint Sample								
	Output Prices		Output (m^2)		Hours Worked		Warp Thread Ball (kg)	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.49 ***		-0.26 ***		0.05 **		0.15 ***	
	(.09)		(.09)		(.02)		(.05)	
Takeup		0.85 ***		-0.47 ***		0.08 **		0.25 ***
		(.16)		(.17)		(.04)		(.08)
R-squared	0.25	0.24	0.24	0.22	0.12	0.13	0.24	0.24
Observations	666	666	676	676	678	678	600	600

Notes: Table reports treatment effects on firm prices, output, hours worked and size of the warp thread ball, all measured in logs. The TOT regression instruments takeup with treatment. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 8: Impacts on Input Prices and Quantities

	Weft Thread Price		Warp Thread Price		Weft Thread Quantity		Warp Thread Quantity	
Panel A: Sample 2								
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.23 ***		-0.03		-0.13		0.08	
	(.04)		(.03)		(.09)		(.09)	
Takeup		0.29 ***		-0.04		-0.16		0.10
		(.05)		(.04)		(.10)		(.11)
R-squared	0.12	0.13	0.13	0.13	0.17	0.17	0.14	0.14
Observations	376	376	376	376	375	375	375	375
Panel B: Joint Sample								
	Weft Thread Price		Warp Thread Price		Weft Thread Quantity		Warp Thread Quantity	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Treatment	0.20 ***		-0.04		-0.19 **		0.08	
	(.06)		(.03)		(.10)		(.09)	
Takeup		0.33 ***		-0.07		-0.34 **		0.10
		(.10)		(.06)		(.17)		(.11)
R-squared	0.23	0.25	0.24	0.25	0.23	0.22	0.14	0.14
Observations	564	564	685	685	677	677	375	375

Notes: Table reports treatment effects on input price and input quantities, all measured in logs. The TOT regression instruments takeup with treatment. The regressions control for baseline values of the dependent variable, and include round and strata fixed effects. Standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 9: Impact of Exporting on Quality Levels

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

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

Notes: Panel A stacks the quality metrics and interacts treatment (ITT) or takeup (TOT) with a quality metric indicator, so each coefficient is the differential impact for each metric between treatment and control. The TOT instruments takeup (interacted with quality metric) with treatment (also interacted with quality metric). Each regression includes baseline values of the quality metric, strata and round fixed effects, and each of these controls is interacted with quality metric indicators. 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 10: Impact of Exporting on Productivity

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

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

Notes: Table reports treatment effects on the two productivity measures: (log) output per hour and total factor productivity. 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 11: Conditional Quality and Productivity

Panel A: Stacked Quality Metrics

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

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

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

Table 12: Specification-Adjusted Quality and Productivity

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

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

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

Table 13: Quality and Productivity on Identical Domestic Rugs

Panel A: Quality Metrics

	Sample 2				Joint Sample			
	Master Artisan		Professor		Master Artisan		Professor	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Packedness	1.09 ***	1.19 ***	0.43 ***	0.48 ***	0.77 ***	1.10 ***	0.28 **	0.43 ***
	(0.15)	(0.15)	(0.14)	(0.15)	(0.13)	(0.16)	(0.11)	(0.16)
Corners	0.89 ***	0.96 ***	0.40 **	0.42 **	0.72 ***	1.05 ***	0.29 **	0.45 **
	(0.16)	(0.15)	(0.17)	(0.18)	(0.14)	(0.17)	(0.13)	(0.18)
Waviness	0.72 ***	0.78 ***	0.29 *	0.30 *	0.55 ***	0.83 ***	0.25 **	0.36 **
	(0.17)	(0.16)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
Weight	0.85 ***	0.96 ***	0.62 **	0.74 ***	0.62 ***	0.91 ***	0.58 ***	1.01 ***
	(0.16)	(0.16)	(0.24)	(0.26)	(0.13)	(0.17)	(0.17)	(0.27)
Touch	0.69 ***	0.77 ***	0.43 ***	0.47 ***	0.50 ***	0.76 ***	0.36 ***	0.53 ***
	(0.13)	(0.13)	(0.15)	(0.16)	(0.11)	(0.15)	(0.12)	(0.17)
Warp Thread Tightness	0.66 ***	0.71 ***	0.43 ***	0.49 ***	0.51 ***	0.74 ***	0.25 **	0.39 **
	(0.10)	(0.10)	(0.15)	(0.15)	(0.09)	(0.12)	(0.12)	(0.17)
Firmness	1.04 ***	1.13 ***	0.44 ***	0.49 ***	0.71 ***	1.01 ***	0.29 **	0.43 **
	(0.14)	(0.14)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
Design Accuracy	0.68 ***	0.79 ***	0.45 ***	0.48 ***	0.53 ***	0.83 ***	0.27 **	0.39 **
	(0.14)	(0.15)	(0.12)	(0.14)	(0.11)	(0.16)	(0.11)	(0.16)
Warp Thread Packedness	1.12 ***	1.20 ***	0.57 ***	0.65 ***	0.87 ***	1.28 ***	0.39 ***	0.62 ***
	(0.16)	(0.16)	(0.15)	(0.16)	(0.14)	(0.18)	(0.12)	(0.17)
R-squared	0.31	0.34	0.10	0.08	0.21	0.32	0.11	0.11
Observations	1,086	1,086	1,078	1,078	1,679	1,679	1,667	1,667

Panel B: Stacked Quality Metrics

	Sample 2				Joint Sample			
	Master Artisan		Professor		Master Artisan		Professor	
	ITT	TOT	ITT	TOT	ITT	TOT	ITT	TOT
	(1A)	(1B)	(2A)	(2B)	(3A)	(3B)	(4A)	(4B)
Stacked Quality Metric	0.86 ***	0.95 ***	0.45 ***	0.50 ***	0.64 ***	0.94 ***	0.33 ***	0.48 ***
	(0.12)	(0.11)	(0.12)	(0.13)	(0.10)	(0.12)	(0.10)	(0.13)
R-squared	0.29	0.34	0.09	0.09	0.19	0.32	0.09	0.13
Observations	1,086	1,086	1,078	1,078	1,679	1,679	1,667	1,667

Panel C: Additional Quality Metrics

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

Notes: Table reports ITT and TOT specifications using the 11 quality metrics from the quality lab. For Panel A, the ITT reports the interaction of the quality metric with treatment dummy, and the TOT reports the interaction of the quality metric with takeup, where takeup is instrumented with quality metric interacted with treatment. Panel B reports the results when the metrics are stacked. Columns 1 and 3 report scores from the master artisan. Columns 2 and 4 report scores from the Professor of Handicraft Science. Panel C reports 3 additional quality metrics and the time spent to produce the rug. All regressions include interactions of strata fixed effects with quality metric, and standard errors are clustered by firm. Significance * .10; ** .05; *** .01.

Table 14: Information Flows and Quality Levels

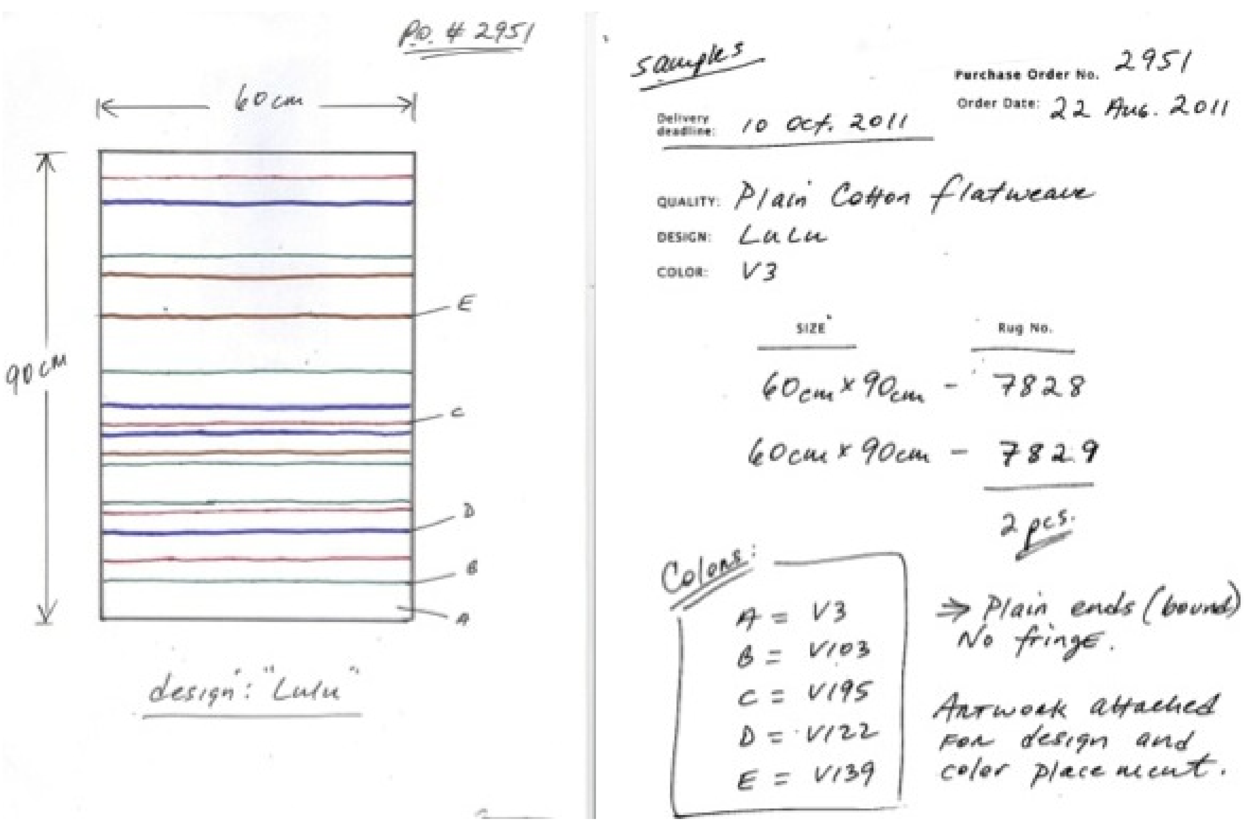
	No Controls		Stacked Quality Metrics Specification Controls		Specification Adjusted	
	Sample 2	Joint Sample	Sample 2	Joint Sample	Sample 2	Joint Sample
	(1)	(2)	(3)	(4)	(5)	(6)
$Takeup_i$ x {Talked About Dimension}$_{id}$	0.19 **	0.19 **	0.16 *	0.18 **	0.16 *	0.16 **
	(0.10)	(0.08)	(0.09)	(0.07)	(0.09)	(0.07)
Quality Metric FEs	yes	yes	yes	yes	yes	yes
Product characteristic controls	no	no	yes	yes	no	no
Specification-adjusted Quality Metrics	no	no	no	no	yes	yes
R-squared	0.75	0.75	0.75	0.75	0.45	0.43
Observations	1,118	1,720	1,088	1,680	1,088	1,687

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

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

Figure 2: Production Technology

Figure 3: Example of Design and Color Patterns Provided by Potential Foreign Client



P.O. # 2951

60 cm

90 cm

E

C

D

B

A

design: "Lulu"

samples

Purchase Order No. 2951

Order Date: 22 Aug. 2011

Delivery deadline: 10 Oct. 2011

QUALITY: Plain Cotton flatweave

DESIGN: Lulu

COLOR: V3

SIZE		Rug No.
60cm x 90cm	-	7828
60cm x 90cm	-	7829
		2 pcs.

Colors:

A = V3
B = V103
C = V195
D = V122
E = V139

→ Plain ends (bound)
No fringe.

Artwork attached for design and color placement.

Figure 4: Example of Rugs Ordered by high-income OECD Clients

Figure 5: Cumulative Export Orders

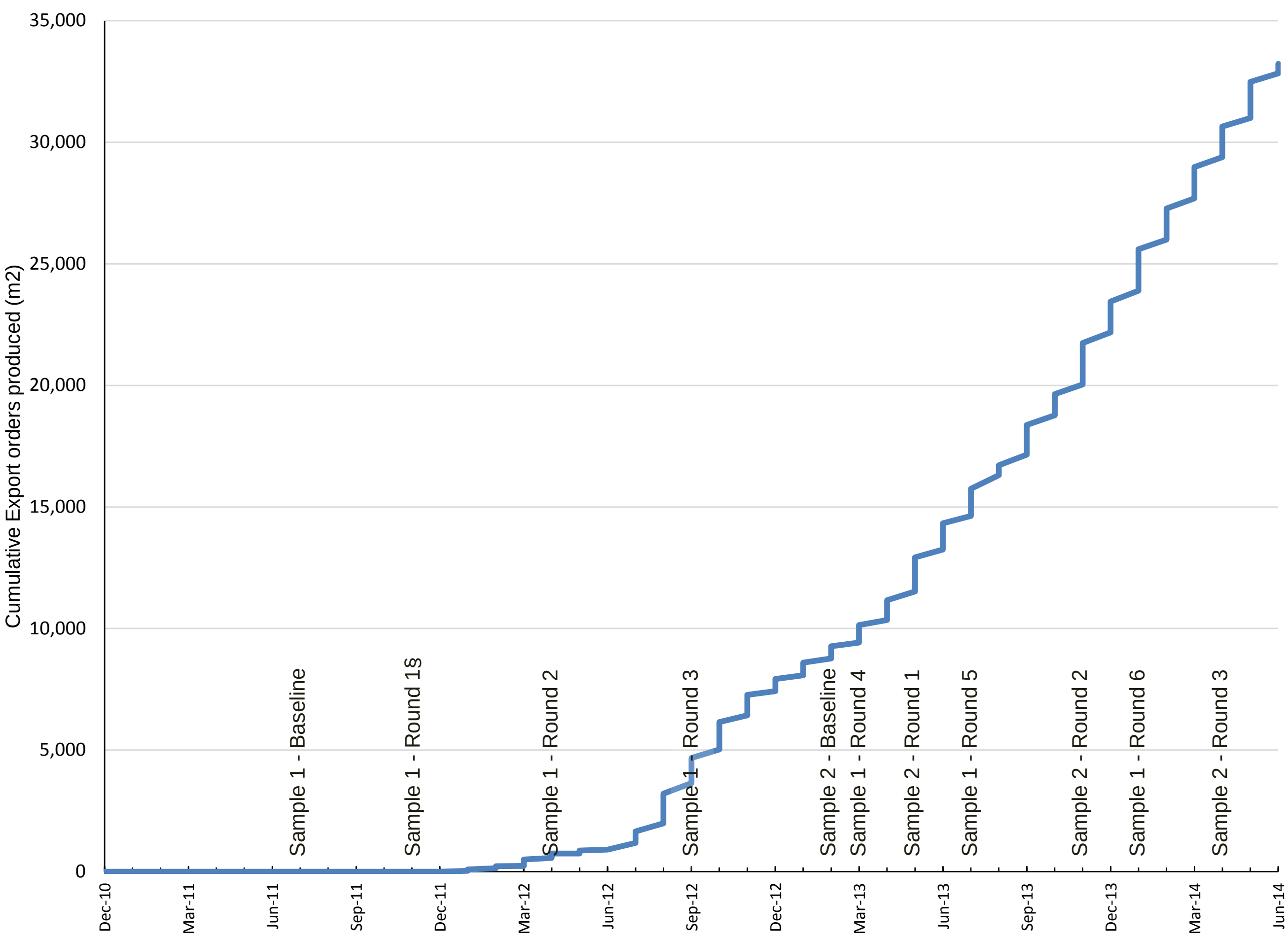


Figure 6: Overlap in Rug Specifications on Domestic and Export Orders

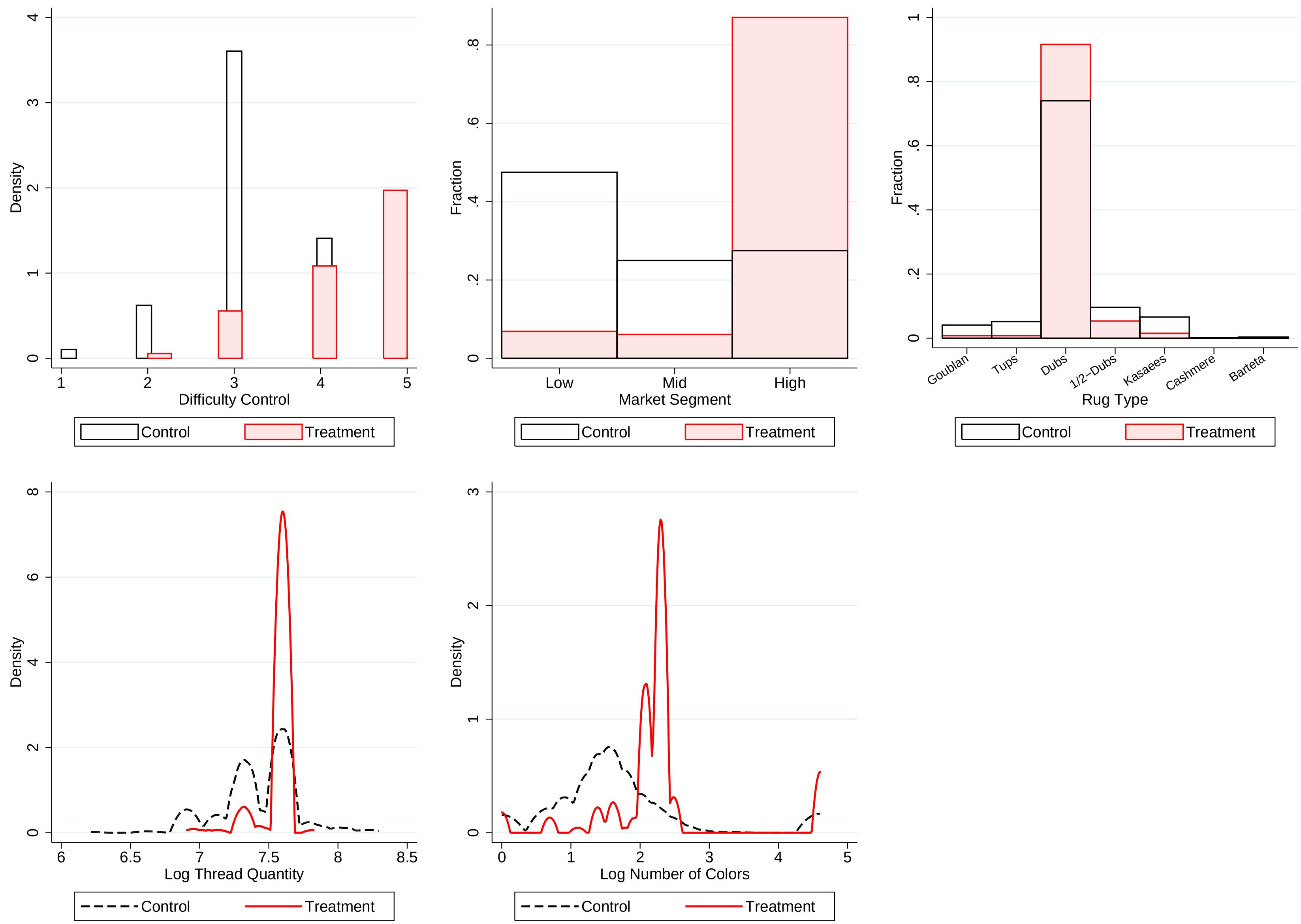


Figure 7: Learning Curves, Sample 2 Takeup Firms

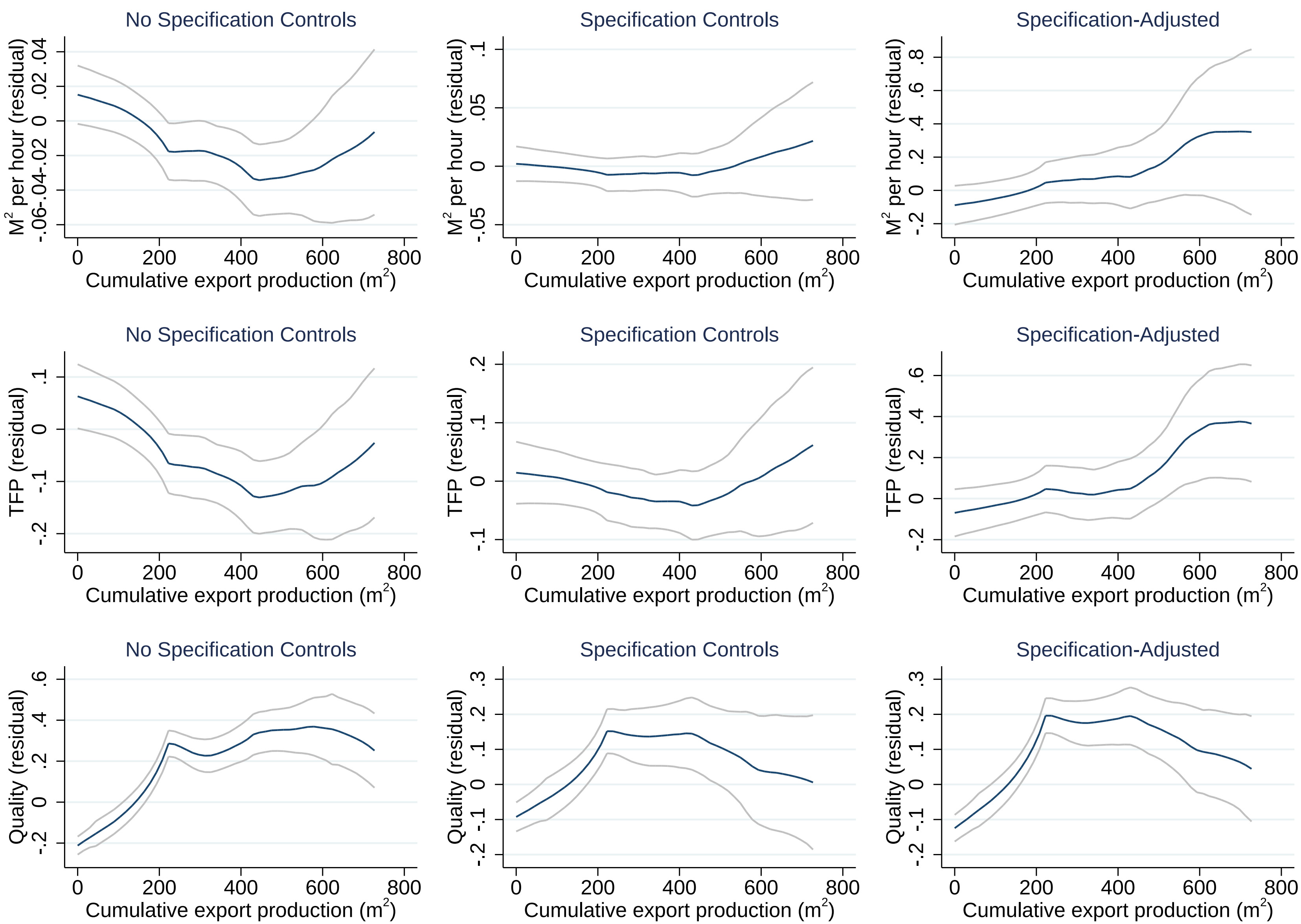
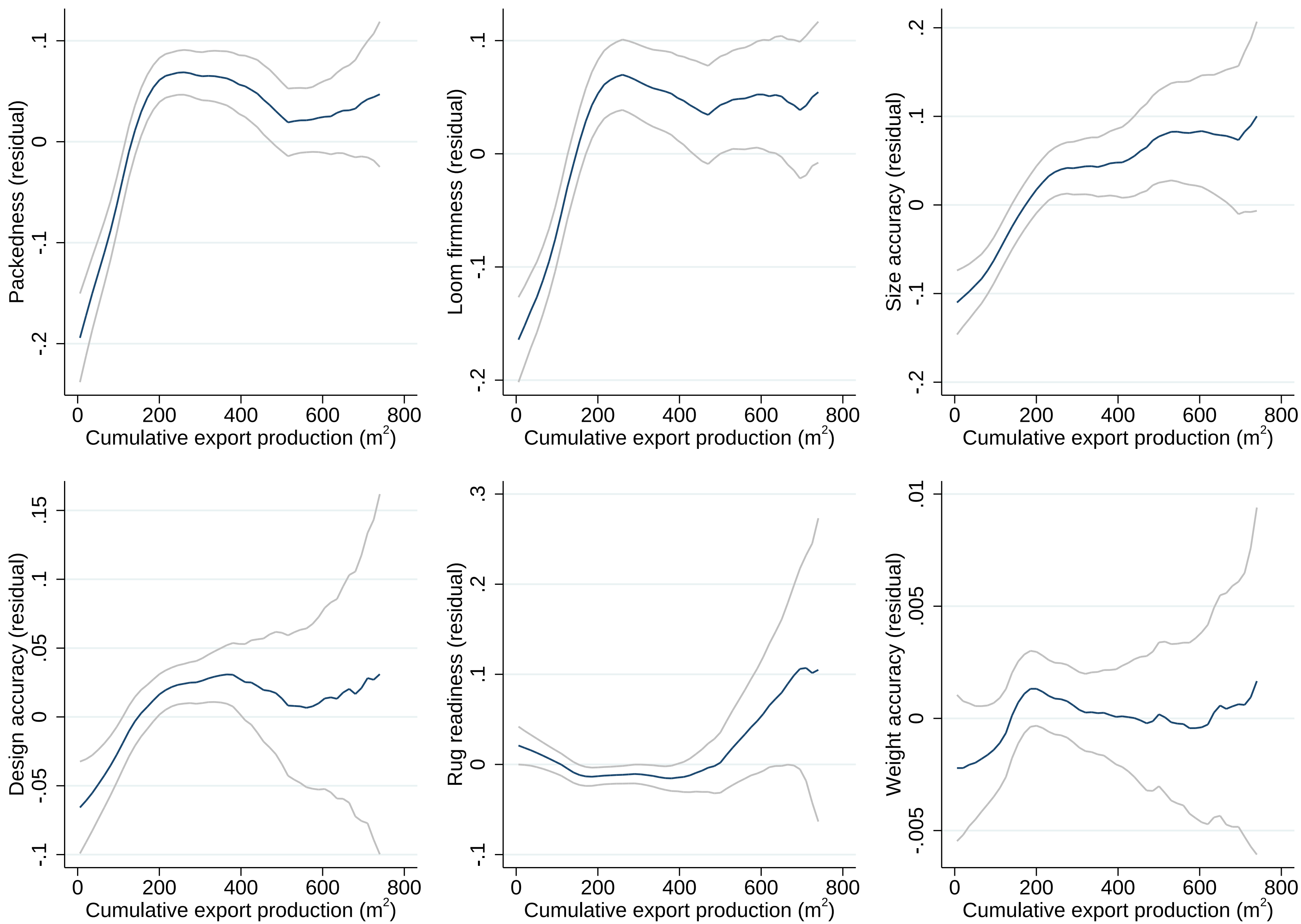


Figure 8: Learning Curves using High-Frequency Order-Book Data, Sample 2 Takeup Firms



Appendix B Tables and Figures not for Publication

Table B.1: Hamis Carpets' Cost Structure

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

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

Table B.2: Conditional and Specification-Adjusted Quality, by Metric

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

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

Table B.3: Summary of Information Flows

	Master Artisan Reported	
	Sample 2	Joint Sample
	(1)	(2)
Number of Visits	10.1	11.0
	(1.76)	(2.57)
Length of Visit (in minutes)	27.8	27.6
	(4.49)	(4.88)
Proportion of firms spoken to about:		
Packedness	0.23	0.20
	(.43)	(.41)
Corners	0.33	0.32
	(.48)	(.47)
Waviness	0.23	0.20
	(.43)	(.41)
Weight	0.50	0.55
	(.51)	(.50)
Touch	0.10	0.11
	(.31)	(.32)
Warp Thread Tightness	0.57	0.48
	(.50)	(.51)
Firmness	0.30	0.32
	(.47)	(.47)
Design Accuracy	0.53	0.50
	(.51)	(.51)
Warp Thread Packedness	0.27	0.23
	(.45)	(.42)
Observations	30	44

Notes: Table summarize the visits between the intermediary. All firms were visited at least 7 times. Length of visit is reported in minutes. The quality measures report whether or not the firm was ever spoken to about that particular quality metric. Standard deviations in parantheses. These data were compiled before the final two take-up firms in Sample 2 began producing for export.

Figure B.1: Learning Curves, Sample 2 Takeup Firms

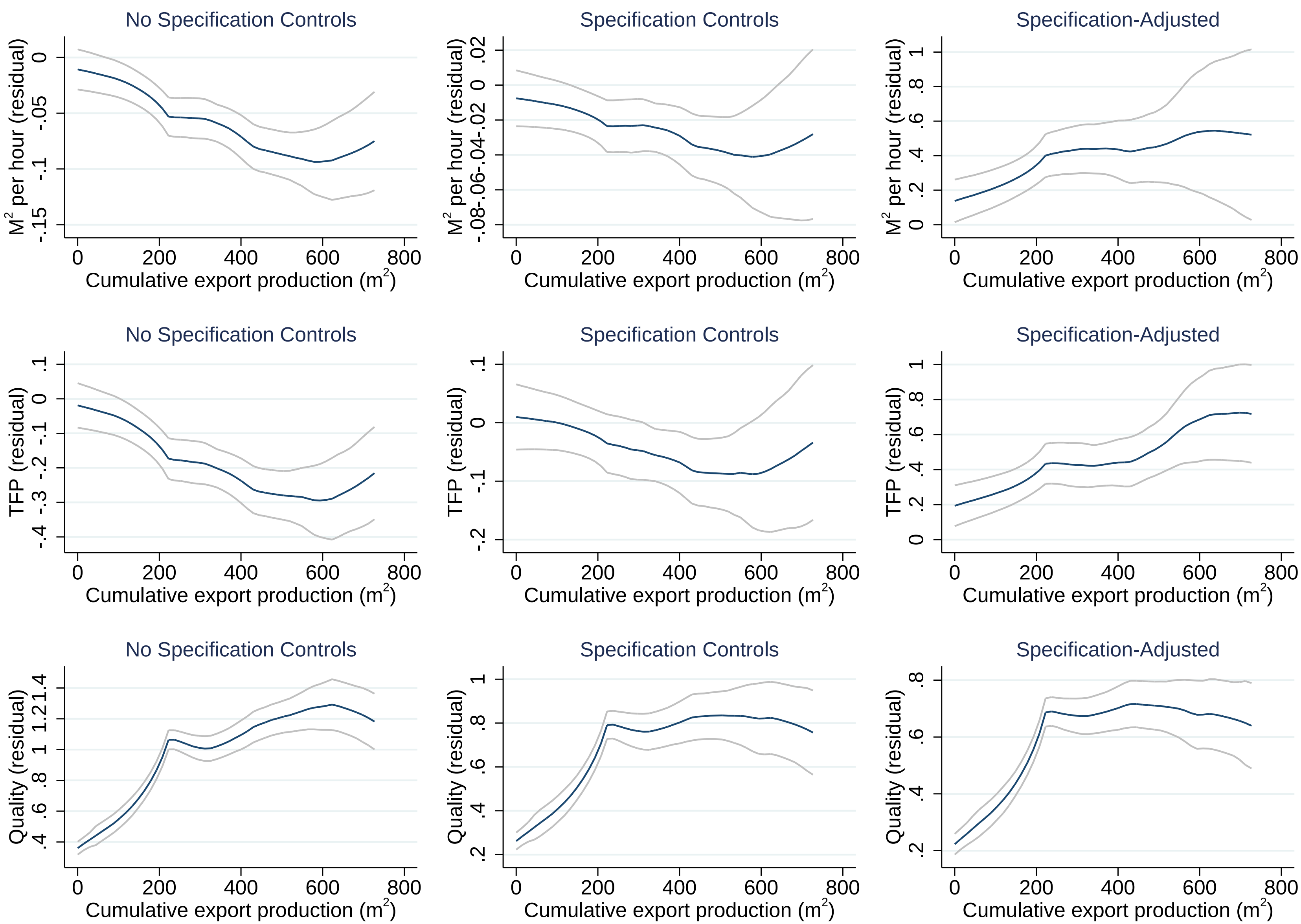


Figure B.2: Learning Curves using High-Frequency Order-Book Data, Sample 2 Takeup Firms

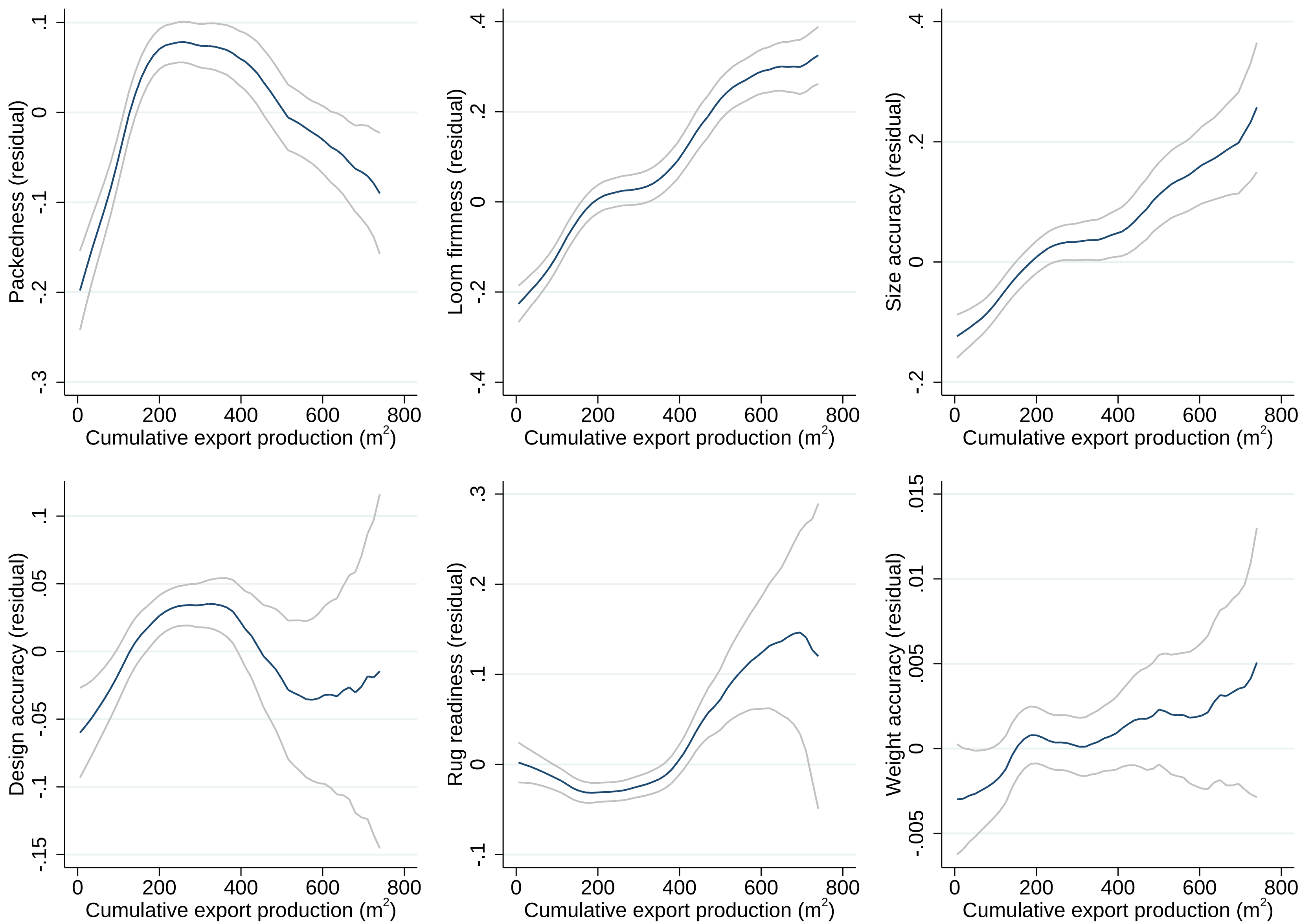


Figure B.3: Quality Problems Noted by Overseas Buyer