DOES MANAGEMENT MATTER? EVIDENCE FROM INDIA

Nicholas Bloom[a], Benn Eifert[b], Aprajit Mahajan[c], David McKenzie[d] and John Roberts[e]

August 9th 2011

Abstract: A long-standing question is whether differences in management practices across firms can explain differences in productivity, especially in developing countries where these spreads appear particularly large. To investigate this we ran a management field experiment on large Indian textile firms. We provided free consulting on management practices to randomly chosen treatment plants and compared their performance to a set of control plants. We find that adopting these management practices raised productivity by 18% through improved quality and efficiency and reduced inventory. Since these practices were profitable this raises the question of why firms had not previously adopted them. Our results suggest that informational barriers were the primary factor explaining this lack of adoption. Since reallocation across firms appeared to be constrained by limits on managerial time, competition did not force badly managed firms to exit.

JEL No. L2, M2, O14, O32, O33.

Keywords: management, organization, productivity and India.

Acknowledgements: Financial support was provided by the Alfred Sloan Foundation; the Freeman Spogli Institute, the International Initiative and the Graduate School of Business at Stanford; the International Growth Centre; IRISS; the Kauffman Foundation; the Murthy Family; the Knowledge for Change Trust Fund; the National Science Foundation; the Toulouse Network for Information Technology; and the World Bank. This research would not have been possible without our partnership with Kay Adams, James Benton and Breck Marshall, the dedicated work of the consulting team of Asif Abbas, Saurabh Bhatnagar, Shaleen Chavda, Karl Gheewalla, Kusha Goyal, Shruti Rangarajan, Jitendra Satpute, Shreyan Sarkar, and Ashutosh Tyagi, and the research support of Troy Smith. We thank our formal discussants Susantu Basu, Francesco Caselli, Ray Fisman, Naushad Forbes, Casey Ichniowski, Vojislov Maksimovic, Ramada Nada, Paul Romer, and Steve Tadelis, as well as seminar audiences at the AEA, Barcelona GSE, Berkeley, BREAD, Boston University, Chicago, Columbia, Cornell, the EBRD, Harvard, Harvard Business School, IESE, Katholieke Universiteit Leuven, Kellogg, the LSE, Maryland, MIT, MIT Sloan, the NBER, NYU, PACDEV, the Sloan Foundation, Stanford, TNIT, Toronto, UBC, UCL, UCLA, UCSC, Victoria, Western Ontario, Wharton, and the World Bank for comments. Nicholas Bloom previously worked as a management consultant at McKinsey and Company, but this consulting firm was not involved in this research experiment.

[a] Stanford Economics, CEP, CEPR, NBER and SIEPR; [b] Overland Advisors LLC, [c] Stanford Economics and SCID; [d] The World Bank, BREAD, CEPR and IZA; [e] Stanford GSB and SIEPR

I. INTRODUCTION

Economists have long puzzled over why there are such astounding differences in productivity across both firms and countries. For example, US plants in industries producing homogeneous goods like cement, block-ice and oak flooring display 100% productivity spreads between the 10th and 90th percentile (Syverson 2004, Foster, Haltiwanger and Syverson, 2008). This productivity dispersion appears even larger in developing countries (Banerjee and Duflo, 2005, Hsieh and Klenow, 2009).

One natural explanation for these productivity differences lies in variations in management practices. Indeed, the idea that "managerial technology" affects the productivity of inputs goes back at least to Walker (1887) and is central to the Lucas (1978) model of firm size. Yet while management has long been emphasized by the media, business schools and policymakers, economists have typically been skeptical about its importance.

One reason for skepticism over the importance of management is the belief that profit maximization will lead firms to minimize costs (e.g. Stigler (1976)). As a result any residual variations in management practices will reflect firms' optimal responses to differing market conditions. For example, firms in developing countries may not adopt quality control systems because wages are so low that repairing defects is cheap. Hence, their management practices are not "bad", but the optimal response to low wages.

A second reason for this skepticism is the complexity of the phenomenon of management, making it hard to measure. Recent work, however, has focused on specific management practices which can be measured, taught in business schools and recommended by consultants. Examples of these practices include key principles of Toyota's "Lean manufacturing", such as quality control procedures, inventory management, and certain human resource management practices. A growing literature measures many such practices and finds large variations across establishments and a strong association between these practices and higher productivity and profitability.[1] However, such correlations may be potentially misleading, for example if profitable firms find it easier to adopt better management practices, thereby accounting for the observed positive correlation.

[1] See for example, the extensive surveys in Lazear and Oyer (2009) and Bloom and Van Reenen (2010). In related work looking at managers (rather than management practices), Bertrand and Schoar (2003) use a manager-firm matched panel and find that manager fixed effects matter for a range of corporate decisions.

This paper provides the first experimental evidence on the importance of management practices in large firms. The experiment takes large, multi-plant Indian textile firms and randomly allocates their plants to treatment and control groups. Treatment plants received five months of extensive management consulting from a large international consulting firm. This consulting diagnosed opportunities for improvement in a set of 38 operational management practices during the first month, followed by four months of intensive support for the implementation of these recommendations. The control plants received only the one month of diagnostic consulting.

The treatment intervention led to significant improvements in quality, inventory and output. We estimate that productivity increased by 18% and annual profitability by about $350,000. Firms also spread these management improvements from their treatment plants to other plants they owned, providing revealed preference evidence on their beneficial impact.

We also find that management practices are more likely to be adopted when productivity is falling, apparently because bad times stimulate the uptake of better management practices. As a result, our fixed effects OLS estimates of the impact of management on productivity are only one third of the experimentally identified IV estimates. This highlights the importance of running field experiments to identify the causal impact of management practices on productivity.

Given this large positive impact of modern management, the natural question is why firms had not previously adopted these practices. Our evidence, while speculative, suggests that informational constraints were the most important factor. For many simple, widespread practices, like the measurement of quality defects, machine downtime and inventory, firms that did not employ them apparently believed that the practices would not improve profits. The owners claimed their quality was as good as other local firms and since they were profitable they did not need to introduce a quality control process. For less common practices, like daily factory meetings, standardized operating procedures, or inventory control norms, firms typically were simply unaware of these practices. While these types of Lean management practices are common in Japan and the US, they are rare in developing countries.

The major challenge of our experiment is its small cross-sectional sample size. We have data on only 28 plants across 17 firms. To address concerns over statistical inference in small samples we implement permutation tests whose size is independent of the sample size. We also exploit our large time series of around 100 weeks of data per plant by using estimators that rely on large T (rather than large N) asymptotics. We believe these approaches are useful for addressing sample concerns in our paper, and also potentially for other field experiments where the data has a small cross-section but long time series.

This paper relates to several strands of literature. First, there is the long literature showing large productivity differences across plants, especially in developing countries. From the outset this literature has attributed much of these spreads to differences in management practices (Mundlak, 1961). But problems in measurement and identification have made this hard to confirm. For example, in Syverson (2011)'s recent survey of the productivity literature he states that "*no potential driving factor of productivity has seen a higher ratio of speculation to empirical study*". Despite this, there are still few experiments on productivity in firms, and none until now involving large multi-plant firms (McKenzie, 2010a).

Second, our paper builds on the literature on firms' management practices. There has been a long debate between the "best-practice" view, that some management practices are universally good so that all firms would benefit from adopting them (Taylor, 1911), and the "contingency view" that optimal practices differ across firms and so observed differences need not reflect bad management (Woodward, 1958). Much of the empirical literature trying to distinguish between these views has been case-study or survey based, making it hard to distinguish between different explanations and resulting in little consensus in the management literature.[2] This paper provides experimental evidence that there are a set of practices that, at least in one industry, would be profitable on average for firms to adopt.

Third, recently a number of other field experiments in developing countries (for example Karlan and Valdivia 2011, Bruhn et al. 2010, Drexler et al. 2010 and Bruhn and Zia 2011) have begun to estimate the impact of basic business training and advice in micro and small enterprises. This research has so far delivered a mix of results. Some studies find significant effects of business training on firm performance while other studies find no effect. The evidence suggests that differences in the quality of the training and the size of the recipient enterprises are important factors determining the impact of this business training. Our research builds on this literature by providing high quality consulting to extremely large multi-plant organizations.

[2] See, for example, the surveys in Delery and Doty (1996), Oyer and Lazear (2009), and Bloom and Van Reenen (2011).

II. MANAGEMENT IN THE INDIAN TEXTILE INDUSTRY

II.A. Why work with firms in the Indian textile industry?

Despite India's recent rapid growth, total factor productivity in India is about 40% of that of the U.S. (Caselli, 2011). While average productivity is low, most notable is the large variation in productivity, with a few highly productive firms and a lot of low-productivity firms (Hsieh and Klenow, 2009).

In common with other developing countries for which data is available, Indian firms are also typically poorly managed. Evidence from this is seen in Figure 1, which plots results from the Bloom and Van Reenen (2010) surveys of manufacturing firms in the US and India. The Bloom and Van Reenen (BVR) methodology scores firms from 1 (worst practice) to 5 (best practice) on management practices related to monitoring, targets, and incentives. Aggregating these scores yields a basic measure of the use of modern management practices that is strongly correlated with a wide range of firm performance measures, including productivity, profitability and growth. The top panel of Figure 1 plots these management practice scores for a sample of 695 randomly chosen US manufacturing firms with 100 to 5000 employees and the second panel for 620 similarly sized Indian ones. The results reveal a thick tail of badly run Indian firms, leading to a lower average management score (2.69 for India versus 3.33 for US firms). Indian firms tend not to collect and analyze data systematically in their factories, they tend not to set and monitor clear targets for performance, and they do not explicitly link pay or promotion with performance. The scores for Brazil and China in the third panel, with an average of 2.67, are similar, suggesting that Indian firms are broadly representative of large firms in emerging economies.

In order to implement a common set of management practices across firms and measure a common set of outcomes, we focus on one industry. We chose textile production since it is the largest manufacturing industry in India, accounting for 22% of manufacturing employment. The fourth panel shows the management scores for the 232 textile firms in the BVR Indian sample, which look very similar to Indian manufacturing in general.

Within textiles, our experiment was carried out on 28 plants operated by 17 firms in the woven cotton fabric industry. These plants weave cotton yarn into cotton fabric for suits, shirts and home furnishings. They purchase yarn from upstream spinning firms and send their fabric to downstream dyeing and processing firms. As shown in the bottom panel of Figure 1, the 17 firms involved had an average BVR management score of 2.60, very similar to the rest of Indian manufacturing. Hence, our particular sample of 17 Indian firms also appears broadly similar in terms of management practices to manufacturing firms in developing countries.

II.B. The selection of firms for the field experiment

The sample firms were randomly chosen from the population of all publicly and privately owned textile firms in Maharashtra, based on lists provided by the Ministry of Corporate Affairs.[3] We restricted attention to firms with between 100 to 1000 employees to focus on larger firms but avoid multinationals. Geographically we focused on firms in the towns of Tarapur and Umbergaon (the largest two textile towns in the area) since this reduced the travel time for the consultants. This yielded a sample of 66 potential subject firms.

All of these 66 firms were then contacted by telephone by our partnering international consulting firm. They offered free consulting, funded by Stanford University and the World Bank, as part of a management research project. We paid for the consulting services to ensure that we controlled the intervention and could provide a homogeneous management treatment to all firms. We were concerned that if the firms made any co-payments they might have tried to direct the consulting, for example asking for help on marketing or finance.

Of this group of firms, 34 expressed an interest in the project and were given a follow-up visit and sent a personally signed letter from Stanford. Of the 34 firms, 17 agreed to commit senior management time to the consulting program.[4] This of courses generates a selection bias in that our results are valid only for the sample of firms that selected into the experiment (Heckman, 1992). To try to evaluate this we compared these program firms with the 49 non-program firms and at least found no significant differences in observables.[5] For example, the program firms had slightly less assets (\$12.8m) compared to the non-program firms (\$13.9m), but this difference was not statistically significant (p-value 0.841). We also compared the groups on management practices using the BVR scores, and found they were almost identical (difference of 0.031, p-value 0.859).

---

[3] The MCA list comes from the Registrar of Business, with whom all public and private firms are legally required to register annually. Of course many firms do not register in India, but this is generally a problem with smaller firms, not with 100+ employee manufacturing firms which are too large and permanent to avoid Government detection.

[4] The main reasons we were given for refusing free consulting were that the firms did not believe they needed management assistance or that it required too much time from their senior management (1 day a week). But it is also possible these firms were suspicious of the offer, given many firms in India have tax and regulatory irregularities.

[5] These observables comprise total assets, employee numbers, total borrowings and the BVR management score.

The experimental firms have typically been in operation for 20 years and all are family-owned. They all produce fabric for the domestic market, and some also export. Table 1 reports some summary statistics for the textile manufacturing parts of these firms (many of the firms have other businesses in textile processing, retail and real estate). On average these firms had about 270 employees, assets of $13 million and sales of $7.5m a year. Compared to US manufacturing firms these firms would be in the top 2% by employment and the top 4% by sales,[6] and compared to India manufacturing in the top 1% by both employment and sales (Hsieh and Klenow, 2010). Hence, these are large manufacturing firms by most standards.[7]

These firms are also complex organizations, with a median of 2 plants per firm (plus a head office in Mumbai) and 4 reporting levels from the shop-floor to the managing director. In all the firms, the managing director is the largest shareholder, and all directors are family members. One firm is publicly quoted on the Mumbai Stock Exchange, although more than 50% of the equity is held by the managing director and his father.

In Exhibits (1) to (7) in the Appendix we include a set of photographs of the plants. These are included to provide some background information to readers on their size, production process and initial state of management. Each plant site involves several multi-story buildings (Exhibit 1). The plants operate a continuous production process that runs day and night (Exhibit 2). The factories' floors were disorganized (Exhibits 3 and 4), and their yarn and spare-parts inventory stores lacked any formalized storage systems (Exhibits 5 and 6), while their quality checking and repair process was extremely labor intensive, employing 19% of workers on average (Exhibit 8).

III. THE MANAGEMENT INTERVENTION

III.A. Why use management consulting as an intervention?

The field experiment aimed to improve management practices in the treatment plants while keeping capital and labor inputs constant. To achieve this we hired a management consultancy firm to work with the plants as the easiest way to rapidly change plant-level management. We selected the consulting firm using an open tender. The winner was a large international management consultancy which is headquartered in the U.S. but has about 40,000 employees in

[6] Dunn & Bradstreet (August 2009) lists 778,000 manufacturing firms in the US with only 17,300 of these (2.2%) with 270 or more employees and only 28,900 (3.7%) with $7.5m or more sales.
[7] Note that most international agencies define large firms as those with more than 250 employees.

India. The full-time team of (up to) 6 consultants working on the project at any time all came from the Mumbai office. These consultants were educated at leading Indian business and engineering schools, and most of them had prior experience working with US and European multinationals.

Selecting a high quality international consulting firm substantially increased the cost of the project.[8] However, it meant that our experimental firms were more prepared to trust the consultants, which was important for getting a representative sample group. It also offered the largest potential to improve the management practices of the firms in our study.

The project ran from August 2008 until August 2010, and the total cost was US$1.3 million, approximately $75,000 per treatment plant and $20,000 per control plant. Note this is very different from what the firms themselves would pay for this consulting, which the consultants indicated would be about $250,000. The reasons for our much lower costs per plant are that the consultancy charged us pro-bono rates (50% of commercial rates) as a research project, provided free partner time and enjoyed considerable economies of scale working across multiple plants.

While the intervention offered high-quality management consulting, the purpose of our study was to use the improvements in management generated by this intervention to understand if (and how) modern management practices affect firm performance. Like many recent development field experiments, this intervention was provided as a mechanism of convenience – to change management practices – and not to evaluate the management consultants themselves.

III.B. The management consulting intervention

The intervention aimed to introduce a set of standard management practices. Based on their prior industry experience, the consultants identified 38 key practices on which to focus. These practices encompass a range of basic manufacturing principles that are standard in almost all US, European and Japanese firms, and can be grouped into five areas:

- Factory Operations: Regular maintenance of machines and recording the reasons for breakdowns to learn from failures. Keeping the factory floor tidy to reduce accidents and ease the movement of materials.

---

[8] At the bottom of the consulting quality distribution in India consultants are cheaper, but their quality is poor. At the top end, rates are similar to those in the US because international consulting companies target multinationals and employ consultants who are often US or European educated and have access to international labor markets.

- Quality control: Recording quality problems by type, analyzing these records daily, and formalizing procedures to address defects to prevent their recurrence.
- Inventory: Recording yarn stocks on a daily basis, with optimal inventory levels defined and stock monitored against these. Yarn sorted, labeled and stored in the warehouse by type and color, and this information logged onto a computer.
- Human-resource management: Performance-based incentive systems for workers and managers. Job descriptions defined for all workers and managers.
- Sales and order management: Tracking production on an order-wise basis to prioritize customer orders by delivery deadline. Using design-wise efficiency analysis so pricing can be based on actual (rather than average) production costs.

These 38 practices (listed in Appendix Table A1) form a set of precisely defined binary indicators that we can use to measure changes in management practices as a result of the consulting intervention.[9] We recorded these indicators on an on-going basis throughout the study. A general pattern at baseline was that plants recorded a variety of information (often in paper sheets), but had no systems in place to monitor these records or use them in decisions. Thus, while 93 percent of the treatment plants recorded quality defects before the intervention, only 29 percent monitored them on a daily basis or by the particular sort of defect, and none of them had any standardized system to analyze and act upon this data.

The consulting treatment had three stages. The first stage, called the *diagnostic* phase, took one month and was given to all treatment and control plants. It involved evaluating the current management practices of each plant and constructing a performance database. Construction of this database involved setting up processes for measuring a range of plant-level metrics – such as output, efficiency, quality, inventory and energy use – on an ongoing basis, plus extracting historical data from existing records. For example, to facilitate quality monitoring on a daily basis, a single metric, termed the Quality Defects Index (QDI), was constructed as a severity-weighted average of the major types of defects. At the end of the diagnostic phase the consulting firm provided each plant with a detailed analysis of its current management practices

[9] We prefer these indicators to the BVR management score for our work here, since they are all binary indicators of specific practices, which are directly linked to the intervention. In contrast, the BVR indicator measures practices at a more general level on a 5-point ordinal scale. Nonetheless, the sum of our 38 pre-intervention management practice scores is correlated with the BVR score at 0.404 (p-value of 0.077) across the 17 firms.

and performance and recommendations for change. This phase involved about 15 days of consulting time per plant over the course of a month.

The second step was a four month *implementation* phase given only to the treatment plants. In this phase, the consulting firm followed up on the diagnostic report to help introduce as many of the 38 key management practices as the firms could be persuaded to adopt. The consultant assigned to each plant worked with the plant management to put the procedures into place, fine-tune them, and stabilize them so that they could readily be carried out by employees. For example, one of the practices was daily meetings for management to review production and quality data. The consultant attended these meetings for the first few weeks to help the managers run them, provided feedback on how to run future meetings, and adjusted their design. This phase also involved about 15 days a month of consulting time per plant.

The third phase was a *measurement* phase, which lasted until August 2010 and involved continued collection of performance and management data from all treatment and control plants. In return for the firms' continuing to provide this data, the consultants provided some light consulting advice to both the treatment and control plants. This phase involved about 1.5 days a month of consulting time per plant.

So, in summary, the control plants were provided with the diagnostic phase and then the measurement phase (totaling 225 consultant hours on average), while the treatment plants were provided with the diagnostic, implementation and then measurement phases (totaling 733 consultant hours on average).

III.C. The experimental design

We wanted to work with large firms because their complexity means systematic management practices are likely to be important. However, providing consulting to large firms is expensive, which necessitated a number of trade-offs detailed below.

Cross-sectional sample size: We worked with 17 firms. We considered hiring cheaper local consultants and providing more limited consulting to a sample of several hundred plants in more locations. But two factors pushed against this. First, many large firms in India are reluctant to let outsiders into their plants because of their lack of compliance with tax, labor and safety regulations. To minimize selection bias we offered a high quality intensive consulting intervention that firms would value enough to take the risk of allowing outsiders into their plants.

This helped maximize initial take-up (26% as noted in section II.B) and retention (100%, as no firms dropped out). Second, the consensus from discussions with members of the Indian business community was that achieving a measurable impact in large firms would require an extended engagement with high-quality consultants. Obviously the trade-off was that this led to a small cross-sectional sample size. We discuss the estimation issues this generates in section III.D below.

Treatment and control plants: The 17 firms in our sample had 28 plants. Due to manpower constraints we could collect detailed performance data from only 20 plants, so we designated 20 plants as "experimental" plants and randomly picked 6 control plants and 14 treatment plants from these. As Table 1 shows, the treatment and control firms were not statistically different across any of the characteristics we could observe.[10] The remaining 8 plants were then the "non-experimental plants": 3 in control firms and 5 in treatment firms. These non-experimental plants did not themselves receive consulting services, but data on their management practices were collected in bi-monthly visits.

Timing: The consulting intervention was executed in three waves because of the capacity constraint of the six-person consulting team. The first wave started in September 2008 with 4 treatment plants. In April 2009 a second wave of 10 treatment plants was initiated, and in July 2009 the diagnostic phase for the 6 control plants was carried out. Firm records usually allowed us to collect data going back to a common starting point of April 2008.

We started with a small first wave because we expected the intervention process to get easier over time due to accumulated experience. The second wave included all the remaining treatment firms because: (i) the consulting interventions take time to affect performance and we wanted the longest time-window to observe the treatment firms; and (ii) we could not mix the treatment and control firms across implementation waves.[11] The third wave contained the control firms. We picked more treatment than control plants because the staggered initiation of the interventions meant the different treatment groups provided some cross identification for each

[10] Treatment and control plants were never in the same firms. The 6 control plants were randomly selected first, and then the 14 treatment plants randomly selected from the remaining 11 firms which did not have a control plant.

[11] Each wave had a one-day kick-off meeting involving presentations from senior partners from the consulting firm. This helped impress the firms with the expertise of the consulting firm and highlighted the potential for performance improvements. Since this meeting involved a project outline, and we did not tell firms about the existence of treatment and control groups, we could not mix the groups in the meetings.

other, and because we believed the treatment plants would be more useful for understanding why firms had not adopted management practices before.

III.D. Small sample size

The focus on large firms meant we had to work with a small sample of firms. This raises three broad issues. A first potential concern is whether the sample size is too small to identify significant impacts. A second is what type of statistical inference is appropriate given the sample size. Third is whether the sample may be too small to be representative of large firms in developing countries. We discuss each concern in turn and the steps we took to address them.

Significance of results: Even though we have only 20 experimental plants across 17 firms, we obtain statistically significant results. There are five reasons for this. First, these are large plants with about 80 looms and about 130 employees each, so that idiosyncratic shocks – like machine breakdowns or worker illness – tend to average out. Second, the data were collected directly from the machine logs, so have very little (if any) measurement error. Third, the firms are homogenous in terms of size, product, region and technology, so that time dummies control for most external shocks. Fourth, we collected weekly data, which provides high-frequency observations over the course of the treatment and the use of these repeated measures can dramatically reduce the sample size needed to detect a given treatment effect (McKenzie, 2010b). Finally, the intervention was intensive, leading to large treatment effects – for example, the point estimate for the reduction in quality defects was almost 50%.

Statistical inference: A second concern is over using statistical tests which rely on asymptotic arguments in the N dimension to justify the normal approximation. We use three alternatives to address this concern. First, we use firm-clustered bootstrap standard errors (Cameron et al, 2008). Second, we implement permutation procedures (for both the Intent to Treat (ITT) and Instrumental Variables estimators) that do not rely upon asymptotic approximations. Third, we exploit our large T sample to implement procedures that rely upon asymptotic approximations along the time dimension (with a fixed N).

Permutation Tests: Permutation procedures use the fact that order statistics are sufficient and complete statistics to propose and derive critical values for test statistics. We first implement this for the null hypothesis of no treatment effect against the two sided alternative for the ITT parameter. This calculates the ITT coefficient for every possible combination of 11 treatment firms out of our 17 total firms (we run this at the firm level to allow for firm-level correlations in errors). Once this is calculated for the 12,376 possible treatment assignments (17 choose 11), the 2.5% and 97.5% confidence intervals are calculated as the 2.5th and 97.5th percentiles of the treatment impact. A treatment effect outside these bounds can be said to be significant at the 5% level. Permutation tests for the IV estimator are more complex, involving implementing a procedure based on Greevy et al. (2004) and Andrews and Marmer (2008) (see Appendix B).

T-asymptotic clustered standard errors: An alternative approach is to use asymptotic estimators that exploit the large time dimension for each firm. To do this we use the recent results by Ibramigov and Mueller (2010) to implement a t-statistic based estimator that is robust to substantial heterogeneity across firms as well as to considerable autocorrelation across observations within a firm. This approach requires estimating the parameter of interest separately for each treatment firm and then treating the resultant set of 11 estimates as a draw from a t distribution with 10 degrees of freedom (see Appendix B). Such a procedure is valid in the sense of having the correct size (for a fixed small number of firms) so long as the time dimension is large enough that the estimate for each firm can be treated as a draw from a normal distribution. In our application we have on average over 100 observations for each firm, so this requirement is likely to be met.

Representativeness of the sample: A third concern with our small sample is how representative it is of large firms in developing countries. In part this concern represents a general issue for field experiments, which are often run on individuals, villages or firms in particular regions or industries. In our situation we focus on one region and one industry, albeit India's commercial hub (Mumbai) and its largest industry (textiles). Comparing our sample to the population of large (100 to 5000 employee) firms in India, both overall and in textiles, suggests that our small sample is at least broadly representative in terms of management practices (see Figure 1). In section V.D we also report results on a plant-by-plant basis to further demonstrate the results are not driven by any particular plant outlier. So while we have a small sample, the results are relatively stable across the individual sample plants.

III.E. The potential conflict of interest in having the consulting firm measuring performance

A final design challenge was the potential for a conflict of interest in having our consulting firm measuring the performance of the experimental firms. To address this we firstly had two graduate students collectively spend six months with the consulting team in India overseeing the daily data collection. Second, about every other month one of the research team visited the firms, meeting with the directors and presenting the quality, inventory and output data the consultants had sent us. This was positioned as a way to initiate discussions on the impact of the experiment with the directors, but it also served to check the data we were receiving reflected reality. If we had presented results showing significant improvements in productivity in firms with no real changes, the directors would have almost certainly challenged the data.

IV. THE IMPACT ON MANAGEMENT PRACTICES

In Figure 2 we plot the average management practice adoption of the 38 practices for the 14 treatment plants, the 6 control plants, and the 8 non-experimental plants. This data is shown at 2 month intervals before and after the diagnostic phase. Data from the diagnostic phase onwards was compiled from direct observation at the factory. Data from before the diagnostic phase was collected from detailed interviews of the plant management team based on any changes to management practices during the prior year. Figure 2 shows five key results:

First, all plants started off with low baseline adoption rates of the set of 38 management practices.[12] Among the 28 individual plants the initial adoption rates varied from a low of 7.9% to a high of 55.3%, so that even the best managed plant in the group used just over half of the key textile-manufacturing practices in place. This is consistent with the results on poor general management practices in Indian firms shown in Figure 1.[13] For example, many of the plants did not have any formalized system for recording or improving production quality, which meant that the same quality defect could arise repeatedly. Most of the plants also had not organized their yarn inventories, so that yarn stores were mixed by color and type, without labeling or computerized entry. The production floor was often blocked by waste, tools and machinery, impeding the flow of workers and materials around the factory.

Second, the intervention did succeed in changing management practices. The treatment plants increased their use of the 38 practices over the period by 37.8 percentage points on average (an increase from 25.6% to 63.4%).

[12] The pre-treatment difference between the treatment, control and other plant groups is not statistically significant, with a p-value on the difference of 0.248 (see Table A1).
[13] Interestingly Clark (1987) suggests Indian textile plants may have even been badly managed in the early 1900s.

Third, not all practices were adopted. The firms arguably adopted the practices that were the easiest to implement and/or had the largest short-run pay-offs, like the daily quality, inventory and efficiency review meetings. If so, the choice of practices was endogenous and it presumably varied with the cost-benefit calculation for each practice.[14]

Fourth, the treatment plants' adoption of management practices occurred gradually. In large part this reflects the time taken for the consulting firm to gain the confidence of the firms' directors. Initially many directors were skeptical about the suggested management changes, and they often started by piloting the easiest changes around quality and inventory in one part of the factory. Once these started to generate improvements, these changes were rolled out and the firms then began introducing the more complex improvements around operations and HR.

Fifth, the control plants, which were given only the 1 month diagnostic, increased their adoption of the management practices, but by only 12% on average. This is substantially less than the increase in adoption in the treatment firms, indicating that the four months of the implementation phase were important in changing management practices.

Finally, the non-experimental plants in the treatment firms also saw a substantial increase in the adoption of management practices. In these 5 plants the adoption rates increased by 17.5%. This increase occurred because the owners of the treatment firms copied the new practices from their experimental plants over to their other plants. Interestingly, this increase in adoption rates is higher even than the control firms' 12% increase, suggesting that the copying of best practices across plants within the same firm can be more effective at improving management practices than short (1-month) bursts of external consulting.

V. THE IMPACT OF MANAGEMENT ON PERFORMANCE

V.A Intention to Treat Estimates

We start by directly estimating the impact of the consulting services which improved management practices via the following intention to treat (ITT) equation:

$$OUTCOME_{i,t} = aTREAT_{i,t} + b_t + c_i + e_{i,t} \quad (1)$$

[14] See Suri (2011) for a related finding on heterogeneous agricultural technology adoption in Kenya.

where OUTCOME is one of the key performance metrics of quality, inventory and output, or total factor productivity (TFP)[15]. TFP is defined as log(value added) – 0.42*log(capital) – 0.58*log(labor), where the factor weights are the cost shares for cotton-weaving in the Indian Annual Survey of Industry (2004-05), capital includes all physical capital (land, buildings, equipment and inventory) and labor is the production hours. $TREAT_{i,t}$ takes the value 1 for the treatment plants after the end of the intervention period and zero otherwise. We exclude data from the six-month period from the start of the diagnostic phase to one month after the end of the implementation phase for both treatment and control firms to ensure a clean differentiation between the before and after periods of the intervention.[16] The b_t are a full set of weekly time dummies to control for seasonality, and the c_i are a full set of plant dummies that are included to control for fixed differences between plants such as the scaling of QDI (per piece, per roll or per meter of fabric) or the loom width (a pick – one pass of the shuttle – on a double-width loom produces twice as much fabric as a pick on single-width loom). The parameter *a* then gives the ITT, which is the average impact of the intervention in the treated plants compared to the control plants.[17]

In Table 2 column (1) we see that the ITT estimate for quality defects is 43.2% which reflects a massive reduction in quality defects.[18] This is shown over time in Figure 3 which plots the Quality Defects Index (QDI) score for the treatment and control plants relative to the start of the treatment period. This is September 2008 for Wave 1 treatment, April 2009 for Wave 2 treatment and control plants.[19] The score is normalized to 100 for both groups of plants using pre-treatment data. To generate point-wise confidence intervals we block-bootstrapped over the firms.

It is clear the treatment plants started to reduce their QDI scores (i.e. improve quality) significantly and rapidly from about week 5 onwards, which was the beginning of the implementation phase following the initial 1 month diagnostic phase. The control firms also

---

[15] We study quality, inventory and output as these easy to measure key production metrics for manufacturing, They also directly influence TFP since poor-quality leads to more mending manpower (increasing labor) and wastes more materials (lowering value-added), high-inventory increases capital, and lower output reduces value-added.

[16] These six months are dropped for both the treatment and control plants. We use a six month window to give us a one month post-intervention margin in the treatment plants.

[17] In the case that the *a* varies across plants, our estimate of *a* will be a consistent estimate of the average value of a_i.

[18] Note that quality is estimated in logs, so that the percentage reduction is 43.2=exp(-0.565)-1.

[19] Since the control plants have no treatment period we set their timing to zero to coincide with the 10 Wave 2 treatment plants. This maximizes the overlap of the data.

showed a mild and delayed downward trend in their QDI scores, consistent with their slower take-up of these practices in the absence of a formal implementation phase.

The reason for this huge reduction in defects is that measuring defects allows firms to address quality problems rapidly. For example, a faulty loom that creates weaving errors would be picked up in the daily QDI score and dealt with in the next day's quality meeting. Without this, the problem would often persist for several weeks, since the checking and mending team had no mechanism (or incentive) for resolving defects. In the longer term the QDI also allows managers to identify the largest sources of quality defects by type, design, yarn, loom and weaver, and start to address these systematically. For example, designs with complex stitching that generate large numbers of quality defects can be dropped from the sales catalogue. This ability to improve quality dramatically through systematic data collection and evaluation is a key element of the lean manufacturing system of production, and in fact many US automotive plants saw reductions in defects of over 90% following the adoption of Lean production systems (see, for example, Womack, Jones and Roos, 1992).

At the foot of table 2 we also present our Ibramigov-Mueller (IM) and permutation significance tests. Both confirm the significance of the reduction in quality defects. First, looking at the IM tests that exploit asymptotics in T rather than N, we find that the IV and ITT results are both significant at the 5% level (zero is outside the 95% confidence intervals). For the standard permutation tests the ITT is again significant at the 5% level (the p-value is 0.04), as are the IV-permutation tests at the 10% level.

Column (2) reports the results for inventory with a 23.9% (=exp(-0.273)-1) post treatment reduction, and Figure 4 shows the plot of inventory over time. The reason for the fall in inventory in treatment plants is that they were carrying about 4 months of raw materials inventory on average before the intervention, including a large amount of dead stock. Because of poor records and storage practices the plant managers typically did not even know they had these stocks. After cataloguing the yarn firms sold or used-up (by incorporation into new designs) the surplus, and then introduced restocking norms for future purchases and monitored against these weekly.

In column (3) we look at output and see a 10.2% (=exp(+0.098)-1) increase in output from the intervention. Several changes drove this increase. First, the halving in quality defects meant the amount of output being scrapped (5% before the beginning of the experiment) fell by about 50%. Second, undertaking routine maintenance of the looms and collecting and monitoring breakdown data helped reduced machine downtime. Visual displays around the factory floor together with the incentive schemes helped to motivate workers to improve operating efficiency and attendance levels. Finally, keeping the factory floor clean and tidy reduced the number of untoward incidents like tools falling into machines or factory fires.

In column (4) show the results for log total factor productivity (TFP) reporting a 18.4% (exp(.169)-1) increase in the treatment firms post treatment compared to the control firms.[20] Productivity increased because output went up (as shown in column (3)), capital dropped (because of lower inventory levels as shown in column (2)) and mending labor dropped (as the number of quality defects fell as shown in column (1)). Figure (5) shows the time profile of productivity.

In columns (5) to (8) we estimate results using a time varying treatment indicator, which is weeks of cumulative implementation. We included these four columns since the changes in management practices and outcomes occurred slowly over the treatment period as Figures 2 to 5 highlight. These results using time since intervention are also all significant on conventional and small-sample robustness test statistics.

We can also examine the difference in quality, inventory and output after treatment on a plant by plant basis. Figure 6 plots the histograms of the before-after changes in our performance measures for the treatment and control plants. It is clear no outliers are driving these differences, with most treatment plants improving on quality (top-left plot), inventory (top-right plot), output (bottom left plot) and productivity (bottom right plot). In comparison the control plants appear to be fairly randomly distributed around the zero impact point, or show small improvements.

Using these results, we can estimate a total increase in profits of around $350,000 per firm, with our calculations outlined in Table A2. We could not obtain accounting data on these firms' profits and losses. Public accounts data are available only with a lag of 2-3 years at the firm level (rather than plant, which is what we would want), and in our interviews with firm owners they told us they under-report profits to avoid tax and also move profits to years when

[20] Note that the 11% increase in productivity reported in the NBER working paper version (Bloom et al. 2011) compares the pre-diagnostic phase to the post-diagnostic phase (including the first five months during the diagnostic and implementation phases). The current numbers exclude the first six months. From Figure 5 it is clear that since the impact is rising during these first six months 'during' treatment dropping this period is appropriate for estimating the full impact of the treatment on productivity. Similarly the profit figures have also increased from $250,000 in the NBER working paper version to $350,000 in the current version since we use the Table 2 ITT estimates.

they want a loan (to have proof of income). When asked for their internal accounts the firms were evasive and would not provide them, beyond occasional comments that profits were in the range of \$0.5m to \$1m per year.[21] So we infer the changes in profits from the improvements in quality, inventory and efficiency. The estimates are medium-run based on the changes over the period of the experiment. In the long-run the impact might be greater if other complementary changes happen (like firms upgrade their design portfolio) or smaller if the firms backslide on these management changes.

To estimate the net increase in profit we also need to calculate the *direct* costs of implementing these changes (ignoring for now any costs of consulting). These costs were small, averaging less than \$3,000 per firm.[22] So given the \$250,000 this consulting would have cost these firms if they paid directly, this implies about a 140% one-year rate of return.

V.B Are the improvements in performance due to Hawthorne effects?

Hawthorne effects are named after a series experiments carried out at the General Electric Hawthorne Works in the 1920s and 1930s. The results apparently showed that just running experiments and collecting data can improve performance, raising concerns that our results could be spurious.

However, we think this is unlikely, for a number of reasons. First, our control plants also had the consultants on site over a similar period of time as the treatment firms. Both sets of plants got the initial diagnostic period and the follow-up measurement period, with the only difference being the treatment plants also got an intensive consulting during the intermediate four month implementation stage while the control plants had briefer, but nevertheless frequent, visits from the consultants collecting data. The control plants were not told they were in the control group. Hence, it cannot be simply the presence of the consultants or the measurement of performance that generated the improvement in performance. Second, the improvements in performance took time to arise and they arose in quality, inventory and efficiency, where the majority of the management changes took place. Third, these improvements persisted for many months after the implementation period, so are not some temporary phenomena due to increased attention. Finally, the firms themselves also believed these improvements arose from better

[21] It is not even clear if firms actually keep correct records of their profits given the risk these could find their way to the tax authorities. For example, any employee that discovered these could use these to blackmail the firm.

[22] About \$35 of extra labor to help organize the stock rooms and factory floor, \$200 on plastic display boards, \$200 for extra yarn racking, \$1000 on rewards, and \$1000 for computer equipment.

management practices, which was the motivation for them extensively copying these practices out to their other non-experimental plants (see Figure 2).

V.C Comparing fixed effects and IV estimations of the impact of management on performance

A growing number of papers estimate the impact of management practices on firm and plant performance by running OLS fixed effects regressions of the type[23]:

$$OUTCOME_{i,t} = \alpha_i + \beta_t + \theta MANAGEMENT_{i,t} + \nu_{i,t} \qquad (2)$$

The concern is that changes in management practices are not exogenous to changes in the outcomes that are being assessed, so that the coefficient θ on management could be biased. For example, a firm may start monitoring quality only when it starts to experience a larger than usual number of defects, which would bias the fixed-effect estimate θ downwards. Or firms may start monitoring product quality as part of a major upgrade of workers and equipment, in which case we could misattribute quality improvements from better capital and labor to better management, biasing θ upwards.

Our study provides an opportunity to examine both the fixed effects estimated coefficients with the experimentally identified IV coefficient, enabling an evaluation of the potential bias in fixed effects estimators of the impact of management. To do this we instrument the management practice score with cumulative weeks of the intervention treatment. Specifically, the instrument is 0 for the control firms throughout, and it runs from 1 to 16 during the implementation phase. We use this instrument because of the gradual impact of the consulting intervention on firms' management practices as shown in Figure 2. The exclusion restriction is that the intervention affected the outcome of interest only through its impact on management practices, and not through any other channel. A justification for this assumption is that the consulting firm focused entirely on the 38 management practices in their recommendations to firms, and firms did not hire new labor and made only trivial investments as a result of the intervention during the period of our study.

In table 3 columns (1) and (2) we see the fixed effects estimate for the impact of management practices on quality defects of -0.561 is about one third of the IV estimate of -1.675. One possible reason for this heavy downward bias is measurement error in the

[23] See, for example, Ichniowski et al. (1998), Cappelli and Neumark (2001) and Black and Lynch (2004). The increasing collection of management data – for example the 50,000 establishment US Census 2011 Management and Organization Survey http://bhs.econ.census.gov/bhs/mops/about.html – means this type of analysis will almost certainly become much more common in future.

management variable, causing attention bias in our fixed effects estimates. However, our management practice measures are binary indicators which are collected every other month by the consultants, who are working with these plants on a regular basis, so should be accurately measured. From discussions with the consultants and owners it appears instead that the main reason for this downward bias is that plants were more willing to adopt new management practices when performance was deteriorating compared to when it was stable or improving. This is consistent with a long stream of micro and macro evidence suggesting bad times spur reorganizations (see, for example, Leibenstein 1966, Hall 1981, and Nickell, Nicolitsas and Patterson, 2001). One reason we found was that it was easier for owners and managers to take the time to make management changes in periods of low demand since they were less busy fulfilling orders. A second reason was they were more willing to concede their current management practices were sub-optimal and needed improving when their performance was deteriorating.

In columns (3) to (4) we see a similar downward bias in the fixed effects estimation of the impact of modern management on inventories, and in columns (5) and (6) on the impact on output. Finally, in columns (7) and (8) we report our key TFP numbers and see that the estimated fixed effects coefficient on incremental adoption of management practices on TFP is 0.159 compared to the IV coefficient of 0.488, which is three times higher. The large difference in the fixed effects and IV estimates highlights the importance of running field experiments on management practices and suggests the prior literature may have substantially underestimated the positive impacts of better management on firm performance.

V.D Complementarities and differential impacts among practices

Another interesting question is the extent to which individual management practices are complementary – for example, does the adoption of improved quality management increase the returns to adopting worker incentives (see Milgrom and Roberts, 1990)? A related question is to what extent different practices impact different outcomes - for example, does quality management drive quality improvements? Unfortunately, because we ran only one experiment we cannot identify multiple management practice variables in an IV estimation, and from section V.B we know that the OLS fixed effects estimates can be heavily biased.

However, we can indirectly investigate complementarity by looking at how the practices cluster. If certain bundles of practices always changed together this is supportive of complementarity. Examining the data, however, we find no strong correlations in adoption rates across sub-groups of practices – for example, the primary factor in the principal component analysis of the changes in individual practices loads positively on 35 of the 38 practices (explaining 26% of the variation), while the second factor explains only 13% of the change. This reveals there is not much variation in adoption rates beyond the overall level of adoption (i.e. that some firms responded more proactively to the intervention while others did not). This in part reflects the fact that our intervention aimed to get plants to adopt all 38 practices rather than induce clustering of practices. But nevertheless, the fact that clustering did not occur is not supportive of strong complementarities *within this group of practices*.[24]

VI. WHY DO BADLY MANAGED FIRMS EXIST?

Given the evidence in the prior section of the large impact of modern management practices on productivity and profitability, the obvious question is why these management changes were not introduced before.

VI.A. Why are firms badly managed?

Our experiment does not directly answer this question, but we can use information generated by the experiment and additional information gathered in the field to draw some preliminary conclusions. In particular, we asked the consultants to document (every other month) the reason for the non-adoption of any of the 38 practices in each plant. To do this consistently we developed a flow-chart (Appendix Exhibit 8) which runs through a series of questions to understand the root cause for the non-adoption of each practice. The consultants collected this data from discussions with owners, managers, and workers, plus their own observations.

As an example of how this flow chart works, imagine a plant that does not record quality defects. The consultant would first ask if there was some external constraint, like labor

[24] The level of (rather than change in) adoption has a similar lack of clustering on sub-groups of practices. For example, looking at the final adoption rate the primary factor loaded positively on all but two of the practices (explaining 29% of the variation), while the second factor explained only 11% of the variation. Of course complementarity is likely to be important for many other bundles of practices (e.g. Ichniowski et al. 1998).

regulations, preventing this, which we found never to be the case.[25] They would then ask if the plant was aware of this practice, which in the example of recording quality typically was the case. The consultants would then check if the plant could adopt the practice with the current staff and equipment, which again for quality recording systems was always true. Then, they would ask if the owner believed it would be profitable to record quality defects, which was often the constraint on adopting this practice. The owner frequently argued that quality was so good they did not need to record quality defects. This view was mistaken, however, because, while these plants' quality might have been good compared to other low-quality Indian textile plants, it was very poor by international standards. So, in this case the reason for non-adoption would be "incorrect information" as the owner appeared to have incorrect information on the cost-benefit calculation.

The overall results for non-adoption of management practices are tabulated in Table 4, for the treatment plants and control plants. This is tabulated at two-month intervals starting the month before the intervention. The rows report the different reasons for non-adoption as a percentage of all practices. These are split into non-adoption reasons for *common* practices (those that 50% or more of the plants were adopting before the experiment like quality and inventory recording or worker bonuses) and *uncommon* practices (those that less than 5% of the plants were adopting in advance like quality and inventory review meetings or manager bonuses). From the table several results are apparent:

First, for the common practices the major initial barrier to adoption was that firms had heard of the practices but thought they would not be profitable to adopt. For example, many of the firms were aware of preventive maintenance but few of them thought it was worth doing. They preferred to keep their machines in operation until they broke down, and then repair them. This accounted for slightly over 45% of the initial non-adoption of practices.

Second, for the uncommon practices the major initial barrier to the adoption was a lack of information about their existence. Firms were simply not aware of these practices. These practices included daily quality, efficiency and inventory review meetings, posting standard-operating procedures and having visual aids around the factory. Many of these are derived from

[25] This does not mean labor regulations do not matter for some practices – for example firing underperforming employees – but they did not directly impinge adopt the immediate adoption of the 38 practices.

the Japanese-inspired Lean manufacturing revolution and are now standard across the US, Japan and Northern Europe but not in developing countries.[26]

Third, as the intervention progressed the lack of information constraint was rapidly overcome in both treatment and control firms. It was easy to explain the existence of these uncommon management practices. This meant that the non-adoption rates of these practices fell relatively rapidly: from 98.5% in the treatment groups one month before the experiment to 63.2% at nine months (a drop of 35.3%).

Fourth, the incorrect information constraints were harder to address. This was because the owners often had strong prior beliefs about the efficacy of a practice and it took time to change these. This was often done using pilot changes on a few machines in the plant or with evidence from other plants in the experiment. For example, the consultants typically started by persuading the managers to undertake preventive maintenance on a set of trial machines, and once it was proven successful it was rolled out to the rest of the factory. And as the consultants demonstrated the positive impact of these initial practice changes, the owners increasingly trusted them and would adopt more of the recommendations, like performance incentives for managers. Thus, the common practice non-adoption rates started at a much lower level but were slower to fall: dropping from 34.6% one month before the experiment for the treatment plants to 16.0% after five months (a drop of 19.4%).

Fifth, changing uncommon practices in the control group was easier than changing common practices. The reason is informing the control group about a new set of practice was easy to do – the consultants simply explained the practice and what this involved. But persuading them to adopt practices which they had already knew but often had strong priors about their lack of efficacy was harder without an active implementation.

Sixth, once the informational constraints were addressed, other constraints arose. For example, even if the owners became convinced of the need to adopt a practice, they would often take several months to adopt it. A major reason is that the owners were severely time constrained, working an average of 68 hours per week already.[27] So, while initially owner's time

[26] This ignorance of best practices seems to be common in many developing contexts, for example in pineapple farming in Ghana (Conley and Udry, 2010).

[27] There was also evidence suggestive of procrastination in that some owners would defer on taking quick decisions for no apparent reason. This matches up with the evidence on procrastination in other contexts, for example African farmers investing in fertilizer (Duflo, Kremer and Robinson, 2011).

accounted for only 3.7% of non-adoption in treatment plants, by 9 months it accounted for 14.0% as a backlog of management changes built up that the owners struggled to implement.

Finally, we did not find evidence for the direct impact of capital constraints, which are a significant obstacle to the expansion of micro-enterprises (e.g. De Mel et al., 2008). Our evidence suggested that these large firms were not cash-constrained, at least for tangible investments. We collected data on all the investments for our 17 firms. The mean (median) investment was $880,000 ($140,000). So investments on the scale of $2000 (the first-year costs of these management changes excluding the consultants' fees) are unlikely to be directly impeded by financial constraints. Of course financial constraints could impede hiring international consultants. The estimated market cost of our free consulting would be $250,000, and as an intangible investment it would be difficult to collateralize. Hence, while financial constraints do not appear to directly block the implantation of better management practices, they may hinder firms' ability to improve their management using external consultants.

VI.B. How do badly managed firms survive?

We have shown that management matters, with improvements in management practices improving plant-level outcomes. One response from economists might then be to argue that poor management can at most be a short-run problem, since in the long run better managed firms should take over the market. Yet, most of our firms have been in business for 20 years and more.

One possible reason why better run firms do not dominate the market may be constraints on growth derived from limited managerial span of control. In every firm in our sample only members of the owning family have positions with major decision-making power over finance, purchasing, operations or employment. Non-family members are given only lower-level managerial positions with authority only over basic day-to-day activities. The principal reason seems to be that family members do not trust non-family members. For example, they are concerned if they let their plant managers procure yarn they may do so at inflated rates from friends and receive kick-backs.[28]

A key reason for this inability to decentralize appears to be the weak rule of law in India. Even if directors found managers stealing, their ability to successfully prosecute them and

[28] This also links to why plant managers did not directly adopt these 38 practices themselves. They had both limited control over factory management and also limited incentives to improve performance since promotion is not possible (only family members can become directors) and there were no bonus systems (the firms did not collect enough performance data).

recover the assets is minimal because of the inefficiency of Indian courts. A compounding reason for the inability to decentralize in Indian firms is the prevalence of bad management practices, as this means the owners cannot keep good track of materials and finance, and so may not even able to identify mismanagement or theft within their firms.[29]

As a result of this inability to delegate, firms can expand beyond the size that can be managed by a single director only if other family members are available to share executive responsibilities. Thus, an important predictor of firm size was the number of male family members of the owners. In particular, the number of brothers and sons of the leading director has a correlation of 0.689 with the total employment of the firm, compared to a correlation between employment and the average management score of 0.223. In fact the best managed firm in our sample had only one (large) production plant, in large part because the owner had no brothers or sons to help run a larger organization. This matches the ideas of the Lucas (1978) span of control model, that there are diminishing returns to how much additional productivity better management technology can generate from a single manager. In the Lucas model, the limits to firm growth restrict the ability of highly productive firms to drive lower productivity ones from the market. In our Indian firms, this span of control restriction seems to be binding, so unproductive firms are able to survive because more productive firms cannot expand.

Entry of new firms into the industry also appears limited by the difficulty of separating ownership from control. The supply of new firms is constrained by the number of families with finance and male family members available to build and run textile plants. Since other industries in India – like software, construction and real estate – are growing rapidly, the attractiveness of new investment in textile manufacturing is relatively limited. Finally, a 50% tariff on fabric imports insulates Indian textile firms against Chinese and other foreign competition.

Hence, the equilibrium appears to be that, with Indian wage rates being extremely low, firms can survive with poor management practices. Because spans of control are constrained, productive firms are limited from expanding, and so reallocation does not drive out badly run firms. And because entry is limited new firms do not enter rapidly. The situation approximates a Melitz (2003) style model with firms experiencing high decreasing returns to scale due to Lucas

[29] A compounding factor is none of these firms had a formalized development or training plan for their managers, so they lacked career motivation. In contrast, Indian software and finance firms that have grown management beyond the founding families place a huge emphasis on development and training (see also Banerjee and Duflo, 2000).

(1978) span of control constraints, high entry costs, and low initial productivity draws (because good management practices are not widespread). The resultant equilibrium has low average productivity, low wages, low average firm-size, and a large dispersion of productivity.[30]

VI.C. Why do firms not use more management consulting?

Finally, why do these firms not hire consultants themselves, given the large gains from better management? A primary reason is that these firms are not aware they are badly managed, as illustrated in Table 4. Of course consulting firms could approach firms for business, pointing out that their practices were bad and offer to fix them. But Indian firms are bombarded with solicitations from businesses offering to save them money on everything from telephone bills to yarn supplies, and so are unlikely to be receptive. Of course consulting firms could go further and offer to provide free advice in return for an *ex post* profit-sharing deal. But monitoring this would be extremely hard, given the firms' desire to conceal profits from the tax authorities. Moreover, the client firm in such an arrangement might worry that the consultant would twist its efforts to increase short-term profits at the expense of long-term profits.

VII. CONCLUSIONS

We implemented a randomized experiment that provided managerial consulting services to textile plants in India. This experiment led to improvements in basic management practices, with plants adopting lean manufacturing techniques that have been standard for decades in the developed world. These improvements in management practices led to improvements in productivity of 18% within the first year from improved quality and efficiency and reduced inventory.

It appears that competition did not drive these badly managed firms out of the market because the inability to delegate decisions away from the owners of the firm impeded the growth of more efficient firms and inter-firm reallocation. Firms had not adopted these management practices before because of informational constraints. In particular, for many of the more widespread practices, while they had heard of these before they were skeptical of their impact. For less common management practices they simply had not heard of the practices.

[30] Caselli and Gennaioli (2011) calibrate an economy with family firms that are unable to grow due to delegation constraints and find a reduction in TFP of 35%, suggesting these kinds of distortions can be quantitatively important.

We also investigated the difference between estimating the impact of management practices on performance using an OLS estimator with firm fixed-effects and instrumenting for management changes using experimental variation. We found the OLS estimate severely downwardly biased, delivering estimates of the impact of management on performance of around one third the IV values. The reason appears to be that firms are more willing to adopt new management practices during periods of deteriorating performance then during periods of stable or improving performance. This highlights the importance of management field experiments for identifying both the sign and the magnitude of the impact of management practices on firm performance.

In terms of future research we would like to investigate the extension of these results to other industries, countries and firm characteristics. In particular, the firms in our experiment are large multi-plant firms operating 24 hours a day across multiple locations, so are complex to manage. Other similarly sized (or larger) firms would presumably also benefit from adopting formalized management practices that continuously monitor the production process. But much smaller firms - such as the typically single-person firms studied in De Mel et al. (2008) – may be simple enough that the owner can directly observe the full production process so does not need formal monitoring systems. Other interesting extensions involve examining the spillover of better management practices across firms within the same industry or region, and the complementarity of different bundles of management practices.

Finally, what are the implications of this for public policy? Certainly we do not want to advocate free consulting, given its extremely high cost. However, our research does support some of the common recommendations to improve productivity, like increasing competition (both from domestic firms and multinationals) and improving the rule of law. More novel is that our results suggest that firms were not implementing best practices on their own because of lack of information and knowledge. This suggests that training programs for basic operations management, like inventory and quality control, could be helpful, as would demonstration projects.

APPENDIX A: DATA

Our estimates for profits are laid out in Table A2, with the methodology outlined below. We calculate the numbers for the median firm. We first generate the estimated impacts on quality, inventory and efficiency using the Intention to Treat (ITT) numbers from Table 2, which shows a reduction of quality defects of 43.2% (exp(-0.565)-1), a reduction in inventory of 23.9% (exp(-0.273)-1) and an increase in output of 10.3% (exp(0.098)-1).

Mending wage bill:

Estimated by recording the total mending hours, which is 71,700 per year on average, times the mending wage bill which is 36 rupees (about $0.72) per hour. Since mending is undertaken on a piece-wise basis – so defects are repaired individually – a reduction in the severity weighted defects should lead to a proportionate reduction in required mending hours.

Fabric revenue loss from non grade-A fabric:

Waste fabric estimated at 5% in the baseline, arising from cutting out defect areas and destroying and/or selling at a discount fabric with unfixable defects. Assume an increase in quality leads to a proportionate reduction in waste fabric, and calculate for the median firm with sales of $6m per year.

Inventory carrying costs:

Total carrying costs of 22% calculated as interest charges of 15% (average prime lending rate of 12% over 2008-2010 plus 3% as firm-size lending premium – see for example http://www.sme.icicibank.com/Business_WCF.aspx?pid), 3% storage costs (rent, electricity, manpower and insurance) and 4% costs for physical depreciation and obsolescence (yarn rots over time and fashions change).

Increased profits from higher output

Increasing output is assumed to lead to an equi-proportionate increase in sales because these firms are small in their output markets, but would also increase variable costs of energy and raw-materials since the machines would be running, and repairs. The average ratio of (energy + raw materials + repairs costs)/sales is 69%, so the profit margin on increased efficiency is 31%.

Labor and capital factor shares:

Labor factor share of 0.58 calculated as total labor costs over total value added using the "wearing apparel" industry in the most recent (2004-05) year of the Indian Annual Survey of industry. Capital factor share defined as 1-labor factor share, based on an assumed constant returns to scale production function and perfectly competitive output markets.

APPENDIX B: ECONOMETRICS

We briefly outline in this section the various econometric procedures we implemented to verify the robustness of our results. We first outline the Ibragimov-Mueller procedure and then briefly discuss the two permutation tests and refer the reader to the original papers for a more detailed discussion.

The proposed procedure by Ibragimov-Mueller (2009) (IM) is useful for our case where the number of entities (firms) is small but the number of observations per entity is large. Their approach can be summarized as follows: Implement the estimation method (OLS, IV, ITT) on each treatment firm separately and obtain 11 firm-specific estimates. Note that we cannot do this for the control firms since there is no within-firm variation for the right hand side for the control firms. Therefore the results from this procedure are essentially based on before-after comparisons for the treatment firms, after using the control firms to remove time period effects.

The procedure requires that the coefficient estimates from each entity are asymptotically independent and Gaussian (but can have different variances). In our case this would be justified by an asymptotics in T argument (recall we have over a 100 observations per plant). In particular, we can be agnostic about the exact structure of correlations between observations within a firm as long as the parameter estimators satisfy a central limit theorem. Subject to this requirement, the extent of correlation across observations within an entity is unrestricted. In addition, different correlation structures across firms are permissible since the procedure allows for different variances for each firm level parameter. This "asymptotic heterogeneity" considerably relaxes the usual assumptions made in standard panel data contexts (such as those underlying the cluster covariance matrices in our main tables). Finally, IM show that the limiting standard Gaussian distribution assumption (for each firm) can be relaxed to accommodate heterogeneous scale mixtures of standard normal distributions as well.

We next summarize the ideas underlying the permutation based tests. We first describe the permutation test for the ITT parameter. We base the test on the Wei-Lachin statistic as described in Greevy et al (2004). The reason for using this statistic is that the permutation test for the IV parameter is a generalization of this procedure and so it is natural to consider this procedure in the first step. Consider the vector of outcomes $\{Y_{i,t}\}_{t=1}^{T}$for plant *i* (we examine each outcome separately). Define the binary random assignment variable for firm *i D_i*. Define the random variable

$$q_{i,j,t} = \mathbb{I}(Z_i > Z_j)\Big(\mathbb{I}(Y_{i,t} > Y_{j,t}) - \mathbb{I}(Y_{i,t} < Y_{j,t})\Big)$$

This variable takes on the values 0, 1 and -1. It is equal to zero if plant i is a control or plant j is a treatment plant and any of the outcome variables for either plant is missing. It is equal to +1 if plant *i* is a treatment plant, plant *j* is a control and the outcome for *i* is larger than the outcome for *j*. It is equal to -1 if plant *i* is a treatment plant, plant *j* is a control and the outcome for *i* is smaller than the outcome for *j*. The Wei-Lachin statistic can be written as

$$T = \sum_{i=1}^{N} Z_i q_i = \sum_{i=1}^{N} Z_i \sum_{t=1}^{T} \sum_{j=1}^{N} q_{i,j,t}$$

Under the null hypothesis of no treatment effect, the treatment outcomes should not be systematically larger than the control outcomes. Specifically, under the null hypothesis and conditional upon the order statistics, each possible candidate value of T has an equal probability of occurring. We use this insight to construct a critical value for the test. Consider one of the $\binom{17}{11}$ combinations of the firm treatment assignment variable *Z*. For each such permutation, compute *T*. Form the empirical distribution of T by considering all possible permutations and record the appropriate quantile for the distribution of *T* thus generated (in the one-sided alternative case this would be the $1-\alpha$ quantile). Finally, reject the null hypothesis of no treatment effect if the original statistic *T* exceeds this quantile.

Greevy et al (2004), show that this test has exact size α for any sample size *n*. Therefore, the conclusions of this test do not rely upon any asymptotic theory. Instead, the results lean heavily on the idea of exchangeability – the property that changing the ordering of a sequence of random variables does not affect their joint distribution. For our application, this notion seems reasonable. Note that exchangeability is weaker than the i.i.d. assumption so for instance outcomes across firms can even be correlated (as long as they are equi-correlated).

Consider next the randomization inference based test for the IV case. We first consider the cross-section. Define the counterfactual model for outcomes $Y_d = \tau + \beta d + \epsilon$ and let D_j denote potential treatment status when treatment assignment is j. Define observed treatment status as $D = ZD_1 + (1-Z)D_0$. In our case, the treatment status is the fraction of the 38 practices that the firm has implemented. The maintained assumption is that the potential outcomes are independent of the instrument *Z* or equivalently (ϵ, D_1, D_0) is independent of *Z* and the error term has mean 0. We observe a random sample on (D, Z, Y_D) and wish to test the null hypothesis $\mathrm{H} : \beta = \beta_0$ against the two-sided alternative. Note that under the null hypothesis, $\tilde{Y} \equiv Y - \tau - \beta_0 D = \epsilon$ is independent of *Z* and we use this fact to construct a test along the lines of the previous test. Consider the analogue of the first equation

$$q_{i,j} = \mathbb{I}(Z_i > Z_j)\left(\mathbb{I}(\tilde{Y}_i > \tilde{Y}_j) - \mathbb{I}(\tilde{Y}_i < \tilde{Y}_j)\right)$$

Where we have replaced the response Y by the response subtracted by $\tau + \beta_0 D$. Note that τ is consistently estimable under the null, so without loss of generality we can treat it as known. For our data, we modify this approach to allow for a panel and covariates (time and plant dummies). This parallels the proposal in Andrews and Marmer (2008) and we can define

$$\tilde{Y_{i,t}} = Y_{i,t} - \beta_0 D_{i,t} - X'_{i,t}\hat{\delta}$$

and we form the statistic as

$$\tilde{T} = \sum_{i=1}^{N} Z_i q_i = \sum_{i=1}^{N} Z_i \sum_{t=1}^{T} \sum_{j=1}^{N} \tilde{q}_{i,j,t}$$

Where

$$\tilde{q}_{i,j,t} = \mathbb{I}(Z_i > Z_j)\left(\mathbb{I}(\tilde{Y}_{i,t} > \tilde{Y}_{j,t}) - \mathbb{I}(\tilde{Y}_{i,t} < \tilde{Y}_{j,t})\right)$$

For each candidate value of β, we form $\{\tilde{Y_{i,t}}\}_{i,t}$ and carry out the permutation test (as described in the ITT case above and noting that we do not use pre-treatment outcomes). We collect the set of values for which we could not reject the null hypothesis (against the two-sided alternative at α=.05) to construct an exact confidence set for β. Although the confidence set constructed in this manner need not be a single interval, in all our estimations, the confidence sets were single intervals.

Table A1: The textile management practices adoption rates

Area	Specific practice	Pre-intervention *level*		Post-intervention *change*	
		Treatment	**Control**	**Treatment**	**Control**
Factory Operations	Preventive maintenance is carried out for the machines	0.429	0.667	0.286	0
	Preventive maintenance is carried out per manufacturer's recommendations	0.071	0	0.071	0.167
	The shop floor is marked clearly for where each machine should be	0.071	0.333	0.214	0.167
	The shop floor is clear of waste and obstacles	0	0.167	0.214	0.167
	Machine downtime is recorded	0.571	0.667	0.357	0
	Machine downtime reasons are monitored daily	0.429	0.167	0.5	0.5
	Machine downtime analyzed at least fortnightly & action plans implemented to try to reduce this	0	0.167	0.714	0
	Daily meetings take place that discuss efficiency with the production team	0	0.167	0.786	0.5
	Written procedures for warping, drawing, weaving & beam gaiting are displayed	0.071	0.167	0.5	0
	Visual aids display daily efficiency loomwise and weaverwise	0.214	0.167	0.643	0.167
	These visual aids are updated on a daily basis	0.143	0	0.643	0.167
	Spares stored in a systematic basis (labeling and demarked locations)	0.143	0	0.143	0.167
	Spares purchases and consumption are recorded and monitored	0.571	0667	0.071	0.167
	Scientific methods are used to define inventory norms for spares	0	0	0.071	0
Quality Control	Quality defects are recorded	0.929	1	0.071	0
	Quality defects are recorded defect wise	0.286	0.167	0.643	0.833
	Quality defects are monitored on a daily basis	0.286	0.167	0.714	0.333
	There is an analysis and action plan based on defects data	0	0	0.714	0.167
	There is a fabric gradation system	0.571	0.667	0.357	0
	The gradation system is well defined	0.500	0.5	0.429	0
	Daily meetings take place that discuss defects and gradation	0.071	0.167	0.786	0.167
	Standard operating procedures are displayed for quality supervisors & checkers	0	0	0.714	0
Inventory Control	Yarn transactions (receipt, issues, returns) are recorded daily	0.929	1	0.071	0
	The closing stock is monitored at least weekly	0.214	0.167	0.571	0.5
	Scientific methods are used to define inventory norms for yarn	0	0	0.083	0
	There is a process for monitoring the aging of yarn stock	0.231	0	0.538	0
	There is a system for using and disposing of old stock	0	0	0.615	0.6
	There is location wise entry maintained for yarn storage	0.357	0	0.357	0
Loom Planning	Advance loom planning is undertaken	0.429	0.833	0.214	0
	There is a regular meeting between sales and operational management	0.429	0.500	0.143	0
Human Resources	There is a reward system for non-managerial staff based on performance	0.571	0.667	0.071	0
	There is a reward system for managerial staff based on performance	0.214	0.167	0.286	0
	There is a reward system for non-managerial staff based on attendance	0.214	0.333	0.357	0
	Top performers among factory staff are publicly identified each month	0.071	0	0.357	0
	Roles & responsibilities are displayed for managers and supervisors	0	0	0.643	0
Sales and Orders	Customers are segmented for order prioritization	0	0	0	0.167
	Orderwise production planning is undertaken	0.692	1	0.231	0
	Historical efficiency data is analyzed for business decisions regarding designs	0	0	0.071	0
All	Average of all practices	0.256	0.288	0.378	0.120
p-value for the difference between the average of all practices		0.510		0.000	

Notes: Reports the 38 individual management practices measured before, during and after the management intervention. The columns **Pre Intervention *level* of Adoption** report the pre-intervention share of plants adopting this practice for the 14 treatment and 6 control plants. The columns **Post Intervention *increase* in Adoption** report the changes in adoption rates between the pre-intervention period and 4 months after the end of the diagnostic phase (so right after the end of the implementation phase for the treatment plants) for the treatment and control plants. The **p-value for the difference between the average of all practices** reports the significance of the difference in the average level of adoption and the increase in adoption between the treatment and control groups.

Table A2: Estimated median impact on profits

Change	Impact	Estimation approach	Estimated impact
Improvement in quality	Reduction in repair manpower	Reduction in defects (43.2%) times median mending manpower wage bill ($41,000).	$18,000
	Reduction in waste fabric	Reduction in defects (43.2%) times the average yearly waste fabric (5%) times median average sales ($6m).	$129,000
Reduction in inventory	Reduction in inventory carrying costs	Reduction in inventory (23.8%) times carrying cost of inventory (22%) times median inventory ($230,000)	$12,000
Increased efficiency	Increased sales	Increase in output (10.3%) times margin on sales (31%) times median sales ($6m)	$192,000
Total			$351,000

Notes: Estimated impact of the improvements in the management intervention on firms' profitability using the ITT estimates in Table 2. Figure calculated for the median firm. See Appendix A for details of calculations for inventory carrying costs, fabric waste, repair manpower and factor shares.

Exhibit 1: Plants are large compounds, often containing several buildings.

Plant entrance with gates and a guard post

Plant surrounded by grounds

Front entrance to the main building

Plant buildings with gates and guard post

Exhibit 2: These factories operate 24 hours a day for 7 days a week producing fabric from yarn, with 4 main stages of production

(1) Winding the yarn thread onto the warp beam

(2) Drawing the warp beam ready for weaving

(3) Weaving the fabric on the weaving loom

(4) Quality checking and repair

Exhibit 3: Many parts of these factories were dirty and unsafe

Garbage outside the factory

Garbage inside a factory

Flammable garbage in a factory

Chemicals without any covering

Exhibit 4: The factory floors were frequently disorganized

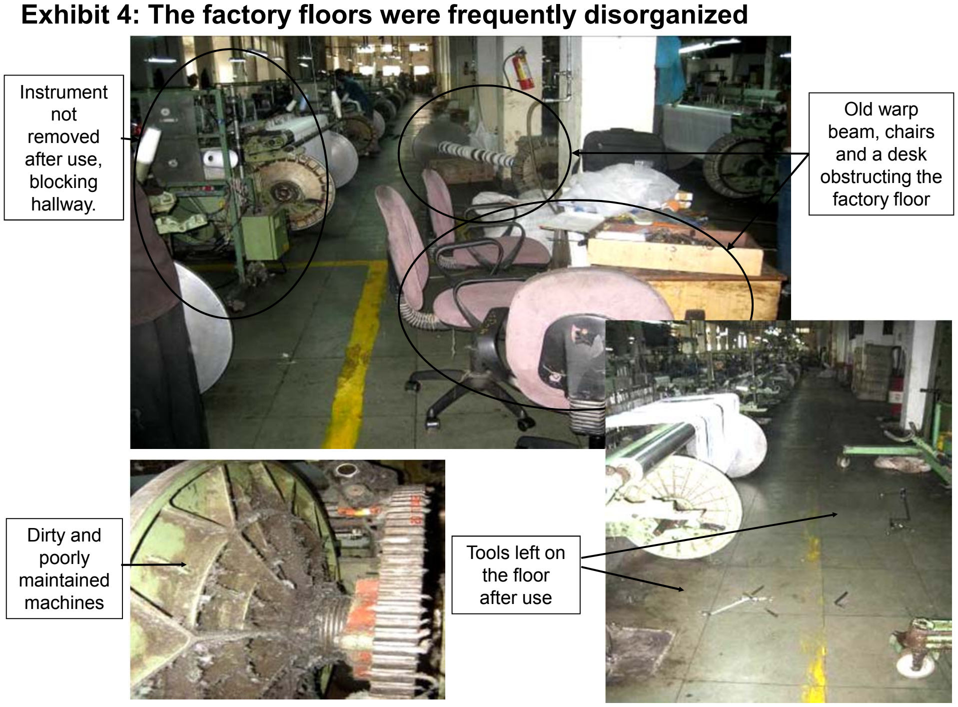


Exhibit 5: Most plants had months of excess yarn, usually spread across multiple locations, often without any rigorous storage system

Exhibit 6: The parts stores were often disorganized and dirty

Spares without any labeling or order

No protection to prevent damage and rust

Spares without any labeling or order

Shelves overfilled and disorganized

Exhibit 7: Poor quality meant that 19% of labor went on repair

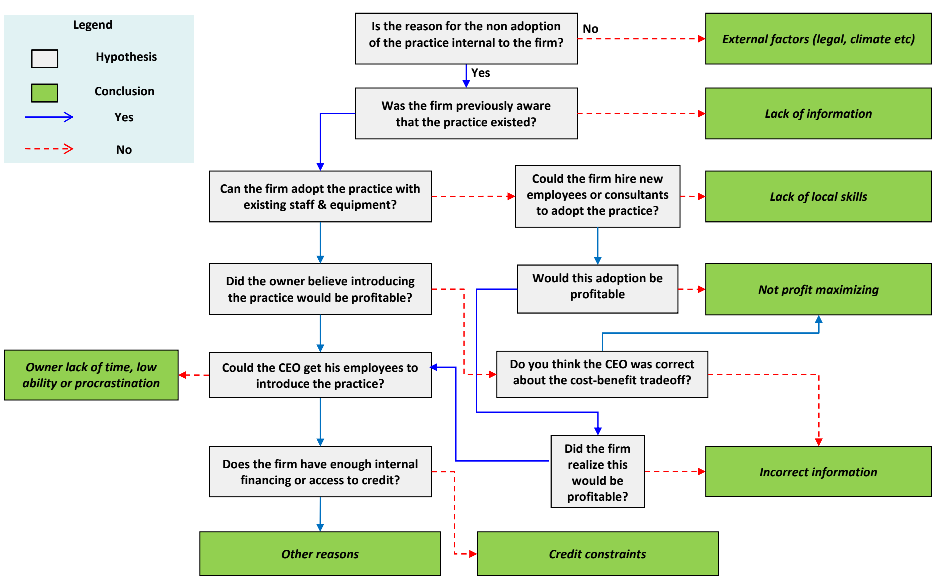
Large room full of repair workers (the day shift)

Workers spread out cloth to spot defects

Defects are repaired by hand or cut out from cloth

Defects lead to about 5% of cloth being scrapped

Exhibit 8: Non adoption flow chart used by consultants to collect data



Notes: The consultants used the flow chart to evaluate why each particular practice from the list of 38 in Table 2 had not been adopted in each firm, on a bi-monthly basis. Non adoption was monitored every other month based on discussions with the firms' directors, managers, workers, plus regular consulting work in the factories.

BIBLIOGRAPHY

Andrews, Donald W.K and Vadim Marmer (2008) "Exactly Distribution-free Inference in Instrumental Variables Regression with Possibly Weak Instruments," *Journal of Econometrics*, 142(1), 183-200

Banerjee, Abhijit and Esther Duflo (2000) "Reputation Effects and the Limits of Contracting: a Study of the Indian Software Industry," *Quarterly Journal of Economics,* 115(3), 989-1017.

Banerjee, Abhijit and Esther Duflo (2005) "Growth Through the Lens of Development Economics", in Philippe Aghion and Stephen Durlauf (eds), *Handbook of Economic Growth, Volume 1,* Amsterdam.

Bertrand, Marianne and Antoinette Schoar (2003) "Managing with Style: The Effects of Managers on Corporate Policy," *Quarterly Journal of Economics,* 118(4), 1169-1208.

Black, Sandra and Lisa Lynch (2004) 'What's Driving the New Economy? The Benefits of Workplace Innovation," *Economic Journal*, 114(493), 97-116.

Bloom, Nicholas, and John Van Reenen (2007) "Measuring and Explaining Management Practices across Firms and Countries", *Quarterly Journal of Economics,* 122(4), 1341-1408.

Bloom, Nicholas, and John Van Reenen (2010) "Why do management practices differ across firms and countries?", *Journal of Economic Perspectives*, March.

Bloom, Nicholas, and John Van Reenen (2011) "Human Resource Management and Productivity," forthcoming *Handbook of Labor Economics.*

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar (2010), "The impact of offering consulting services to small and medium enterprises: evidence from a randomized trial in Mexico," mimeo.

Bruhn, Miriam and Zia, Bilal (2011), "Business and financial literacy for young entrepreneurs: evidence from Bosnia-Herzegovina", World Bank mimeo.

Cameron, A. Colin, Jonah Gelbach and Douglas Miller (2008) "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics* 90(3), 414-27.

Cappelli, P and D. Neumark (2001), "Do 'High Performance" Work Practices Improve Establishment-Level Outcomes?" *Industrial and Labor Relations Review,* 54(4), 737-775.

Caselli, Francesco (2011), "CREI Lectures in Growth", LSE mimeo.

Caselli, Francesco and Gennaioli, Nicolas (2011), "Dynastic management", LSE mimeo.

Clark, Greg (1987), "Why isn't the Whole World Developed?" *Journal of Economic History*, 141-173.

Conley, Tim and Christopher Udry, (2010), "Learning about a New Technology: Pineapple in Ghana," *American Economic Review,* 100(1), 35-69.

De Mel, Suresh, David McKenzie and Christopher Woodruff (2008), "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics,* 113(4), 1329-1372.

Delery, John E. and D. Harold Doty (1996), "Modes of Theorizing in Strategic Human Resource Management: Tests of Universalistic, Contingency, and Configurational Performance Predictions," *Academy of Management Journal,* 39(4), 802-835.

Drexler, Alejandro, Greg Fischer and Antoinette Schoar (2010), "Financial literacy training and rule of thumbs: evidence from a field experiment," mimeo.

Duflo, Esther, Michael Kremer and Jonathan Robinson (2011), "Nudging farmers to use fertilizer: theory and experimental evidence from Kenya," *American Economic Review*, forthcoming.

Foster, Lucia, John Haltiwanger and Chad Syverson (2008) "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98(1), 394-425

Greevy, Robert, Jeffrey H. Silber, Avital Cnaan and Paul R. Rosenbaum (2004), "Randomization Inference with Imperfect Compliance in the ACE-inhibitor," *Journal of the American Statistical Association,* 99(465), 7-15.

Hall, Robert (1991), "Recessions as reorganizations", Stanford mimeo.

Heckman, James (1992), "Randomization in social programs", in C. Manski and I. Garfinkle, eds. *Evaluating Welfare and Training Programs*, Cambridge MA, Harvard University Press.

Hsieh, Chiang-Tai and Pete Klenow (2009), "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics,* 124(4), 1403-1448.

Hsieh, Chiang-Tai and Pete Klenow (2010), "Development accounting," *American Economic Journal: Macroeconomics,* 2(1), 207-223.

Ibragimov, R. and U.K. Muller (2010), "t-statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28(4), 453-468.

Ichniowski, Casey, Kathryn Shaw and Giovanna Prennushi. (1997), "The Effects of Human Resource Management: A Study of Steel Finishing Lines", *American Economic Review*, 87(3), 291-313.

Leibenstein, Harvey (1966) 'Allocative efficiency vs X efficiency' *American Economic Review*, 56: 392–415.

Karlan, Dean and Martin Valdivia (2011) "Teaching Entrepreneurship: Impact of Business Training On Microfinance Clients and Institutions," *Review of Economics and Statistics*, 93(2), 510-27.

Lazear, Edward, and Paul Oyer (2009), "Personnel Economics," in Robert Gibbons and John Roberts, eds. *Handbook of Organizational Economics,* Princeton University Press, forthcoming.

Lucas, Robert (1978), "On the Size Distribution of Business Firms," *Bell Journal of Economics*, 508-23.

McKenzie, David (2010a), "Impact Assessments in Finance and Private Sector Development: What have we learned and what should we learn?" *World Bank Research Observer*, 25(2), 209-33.

McKenzie, David (2010b) "Beyond Baseline and Follow-up: The Case for More T in Experiments," Mimeo, World Bank.

Melitz, Marc (2003), "The impact of Trade on Intra-industry Reallocations and Intra-industry Productivity," *Econometrica,* 71(6), 1695-1725.

Milgrom, Paul and John Roberts (1990), "The Economics of Modern Manufacturing: Technology, Strategy and Organization," *American Economic Review*, 80 (3), 511-528.

Mundlak, Yair (1961), "Empirical Production Function Free of Management Bias," *Journal of Farm Economics*, 43(1), 44–56.

Nickell, Steve, Daphne Nicolitsas and Malcolm Patterson, 2001. 'Does Doing Badly Encourage Innovation?', *Oxford Bulletin of Economics and Statistics*, 63(1), 5-28.

Stigler George, (1976) 'The Xistence of X-efficiency' *American Economic Review*, 66: 213–6.

Syverson, Chad (2004), "Market Structure and Productivity: A Concrete Example", *Journal of Political Economy*, 112(6), 1181-1222.

Syverson, Chad (2011), ''What determines productivity at the micro level?", forthcoming *Journal of Economic Literature*.

Suri, Tavneet (2011), "Selection and comparative advantage in technology adoption", *Econometrica*, 79(1), pp. 159-209.

Taylor, F. (1911), *Principles of Scientific Management*, Harper and Brothers, New York and London.

Walker, Francis (1887), "The source of business profits", *Quarterly Journal of Economics*, 1(3), 265-288.

Womack, James, Daniel Jones and Daniel Roos (1991), *The Machine that Changed the World*, Harper Collins: New York.

Woodward J. (1958), *Management and Technology*, Cambridge: Cambridge University Press.

Table 1: The field experiment sample

	All				Treatment	Control	Diff
	Mean	Median	Min	Max	Mean	Mean	p-value
Sample sizes:							
Number of plants	28	n/a	n/a	n/a	19	9	n/a
Number of experimental plants	20	n/a	n/a	n/a	14	6	n/a
Number of firms	17	n/a	n/a	n/a	11	6	n/a
Plants per firm	1.65	2	1	4	1.73	1.5	0.393
Firm/plant sizes:							
Employees per firm	273	250	70	500	291	236	0.454
Employees, experimental plants	134	132	60	250	144	114	0.161
Hierarchical levels	4.4	4	3	7	4.4	4.4	0.935
Annual sales $m per firm	7.45	6	1.4	15.6	7.06	8.37	0.598
Current assets $m per firm	12.8	7.9	2.85	44.2	13.3	12.0	0.837
Daily mtrs, experimental plants	5560	5130	2260	13000	5,757	5,091	0.602
Management and plant ages:							
BVR Management score	2.60	2.61	1.89	3.28	2.50	2.75	0.203
Management adoption rates	0.262	0.257	0.079	0.553	0.255	0.288	0.575
Age, experimental plant (years)	19.4	16.5	2	46	20.5	16.8	0.662
Performance measures							
Quality defects index	3.89	5.24	0.61	16.4	4.47	7.02	0.390
Inventory (1,000 kilograms)	61.1	72.8	7.4	117.0	61.4	60.2	0.940
Output (picks, million)	23.3	25.4	6.9	32.1	22.1	28.6	0.255
Productivity (in logs)	2.90	2.90	2.12	3.59	2.91	2.86	0.869

Notes: Data provided at the plant and/or firm level depending on availability. **Number of plants** is the total number of textile plants per firm including the non-experimental plants. **Number of experimental plants** is the total number of treatment and control plants. **Number of firms** is the number of treatment and control firms. **Plants per firm** reports the total number of other textiles plants per firm. Several of these firms have other businesses – for example retail units and real-estate arms – which are not included in any of the figures here. **Employees per firm** reports the number of employees across all the textile production plants, the corporate headquarters and sales office. **Employees per experiment plant** reports the number of employees in the experiment plants. **Hierarchical levels** displays the number of reporting levels in the experimental plants – for example a firm with workers reporting to foreman, foreman to operations manager, operations manager to the general manager and general manager to the managing director would have 4 hierarchical levels. **BVR Management score** is the Bloom and Van Reenen (2007) management score for the experiment plants. **Management adoption rates** are the adoption rates of the management practices listed in Table A1 in the experimental plants. **Annual sales ($m)** and **Current assets ($m)** are both in 2009 US $million values, exchanged at 50 rupees = 1 US Dollar. **Daily mtrs, experimental plants** reports the daily meters of fabric woven in the experiment plants. Note that about 3.5 meters is required for a full suit with jacket and trousers, so the mean plant produces enough for about 1600 suits daily. **Age of experimental plant (years)** reports the age of the plant for the experimental plants. **Quality defect index** is a quality weighted measure of production quality defects. **Inventory** is the stock of yarn per intervention. **Output** is the production of fabric in picks (one a pick is single rotation of the weaving shuttle), and **Productivity** which is log(value-added)-0.42*log(capital)-0.58*log(total hours). All performance measures reported pooled across all pre-diagnostic phase data.

Table 2: The impact of modern management practices on plant performance

Dependent Variable	Quality defects	Inventory	Output	TFP	Quality defects	Inventory	Output	TFP
Specification	ITT	ITT	ITT	ITT	Weeks of Treatment	Weeks of Treatment	Weeks of Treatment	Weeks of Treatment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Intervention$_{i,t}$ Post implementation stage	-0.565** (0.231)	-0.273** (0.116)	0.098** (0.036)	0.169** (0.067)				
Cumulative treatment$_{i,t}$ Total weeks of implementation					-0.032** (0.013)	-0.017*** (0.005)	0.006*** (0.002)	0.010** (0.004)
Small sample robustness								
Ibragimov-Mueller (95% CI)	(-0.78,-0.44)	(-0.22,0.01)	(0.22,0.47)	(0.18,0.51)				
(90%CI)	(-0.75,-0.47)	(-0.20,-0.01)	(0.25,0.45)	(0.21,0.48)				
Permutation Test I (p-value)	.04	.13	.04	.05				
Time FEs	125	122	125	122	125	122	125	122
Plant FEs	20	18	20	18	20	18	20	18
Observations	1396	1627	1966	1447	1732	1977	2312	1779

Notes: All regressions use a full set of plant and calendar week dummies. Standard errors bootstrap clustered at the firm level. To compare before and after outcomes the ITT regressions exclude the six months of data from the beginning of the diagnostic phase for all plants (so excluding the diagnostic phase, treatment phase and first month after treatment). **Quality defects** is a log of the quality defects index (QDI), which is a weighted average score of quality defects, so higher numbers imply worse quality products (more quality defects). **Inventory** is the log of the tons of yarn inventory in the plant. **Output** is the log of the weaving production picks. **TFP** is plant level total factor productivity defined as log(output) measured in production picks less log(capital) times capital share of 0.42 less log(labor) times labor costs share of 0.58. **Intervention** is a plant level indicator taking a value of 1 after the implementation phase at treatment plants and zero otherwise. **Cumulative treatment** is the cumulative weeks of since beginning the implementation phase in each plant (zero in the control groups and prior to the implementation phase). **ITT** reports the intention to treat results from regressing the dependent variable directly on the 1/0 intervention indicator. **Time FEs** report the number of calendar week time fixed effects. **Plant FEs** reports the number of plant-level fixed effects. Two plants do not have any inventory on site, so no inventory data is available. **Small sample robustness** implements three different procedures (described in greater detail in Appendix B) to address issues of plant heterogeneity, within plant (and firm) correlation, and small sample concerns, where **95% CI** and **90% CI** report 95% and 90% confidence intervals. **Ibragimov-Mueller** estimates parameters firm-by-firm and then treats the estimates as draws from independent (but not identically distributed) normal distributions. **Permutation Test I** reports the p-values for testing the null hypothesis that the treatment has no effect for the ITT parameter by constructing a permutation distribution of the ITT estimate using the 12376 possible permutations of treatment assignment. These tests have exact finite sample size.

Table 3: OLS and IV estimations of the impact of management practices on plant performance

Specification	OLS	IV 2nd stage	OLS	IV 2nd stage	OLS	IV 2nd stage	OLS	IV 2nd stage
Dependent Variable	Quality defects	Quality defects	Inventory	Inventory	Output	Output	TFP	TFP
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Management}_{i,t}$	-0.561	-1.675**	-0.639**	-0.921***	0.127	0.320***	0.159	0.488**
Adoption of management practices	(0.450)	(0.763)	(0.253)	(0.290)	(0.103)	(0.118)	(0.178)	(0.227)
Specification		IV 1st stage		IV 1st stage		IV 1st stage		IV 1st stage
Dependent Variable:		Management		Management		Management		Management
$\text{Cumulative treatment}_{i,t}$		0.019***		0.018***		0.019***		0.020***
Total weeks of implementation		(0.002)		(0.002)		(0.002)		(0.002)
Small sample robustness								
Ibragimov-Mueller (95% CI)	(-2.21,-0.91)	(-2.63,-1.22)	(-0.81,0.03)	(-0.70,0.00)	(-2.48.,2.46)	(0.52,1.71)	(-3.42,2.93)	(0.20,2.12)
(90%CI)	(-2.09,-1.03)	(-2.50,-1.36)	(-0.73,,-0.06)	(-0.64,-0.05)	(-2.02,2.00)	(0.63,1.59)	(-2.82,2.35)	(0.39,1.93)
IV Permutation Tests (95% CI)		(-4.05,-.06)		(-8.29,1.18)		(0.14,1.11)		(0.40,10.5)
(90% CI)		(-4.00,-0.43)		(-5.21,0.65)		(0.19,0.94)		(0.56,3.15)
First stage F-test		67.51		63.76		91.20		74.68
Time FEs	113	113	113	113	114	114	113	113
Plant FEs	20	20	18	18	20	20	20	20
Observations	1732	1732	1977	1977	2312	2312	1779	1779

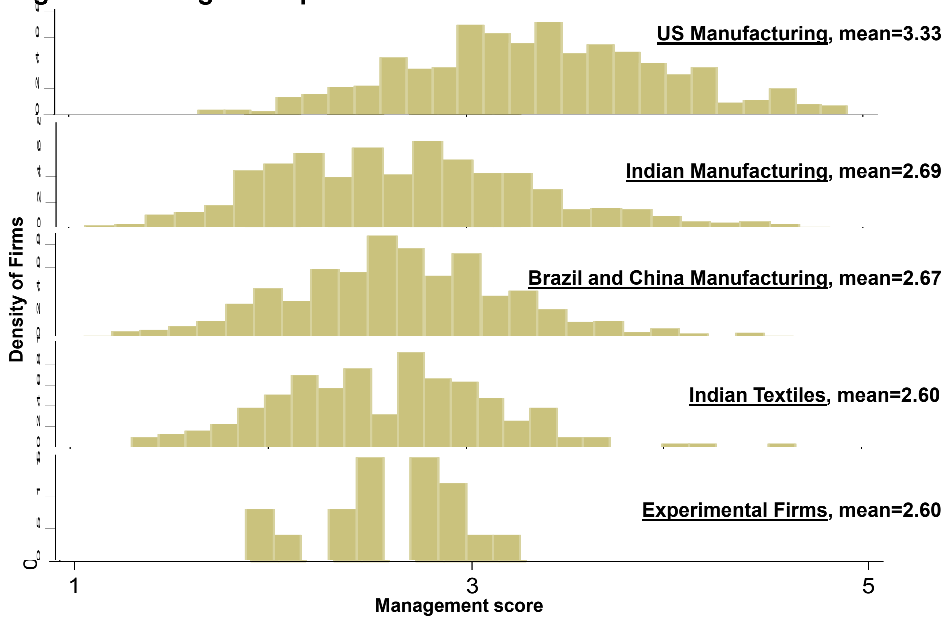
Notes: All regressions use a full set of plant and calendar week dummies. Standard errors bootstrap clustered at the firm level. **Quality defects** is a log of the quality defects index (QDI), **Inventory** is the log of the tons of yarn inventory in the plant. **Output** is the log of the weaving production picks. **Management** is the adoption share of the 38 management practices listed in table A1. **Cumulative treatment** is the cumulative weeks of since beginning the implementation phase in each plant (zero in the control groups and prior to the implementation phase). **OLS** reports results with plant estimations. **IV** reports the results where the management variable has been instrumented with weeks of cumulative treatment. **Time FEs** report the number of calendar week time fixed effects. **Plant FEs** reports the number of plant-level fixed effects. Two plants do not have any inventory on site, so no inventory data is available. **Small sample robustness** implements three different procedures (described in greater detail in Appendix B) to address issues of plant heterogeneity, within plant (and firm) correlation, and small sample concerns, where **95% CI** and **90% CI** report 95% and 90% confidence intervals. **Ibragimov-Mueller** estimates parameters firm-by-firm and then treats the estimates as a draw from independent (but not identically distributed) normal distributions. **Permutation Test I** reports the p-values for testing the null hypothesis that the treatment has no effect for the ITT parameter by constructing a permutation distribution of the ITT estimate using 1000 possible permutations (out of 12376) of treatment assignment. **IV-Permutation** tests implements a permutation test for the IV parameter using 1000 possible permutations (out of 12376) of treatment assignment. These tests have exact finite sample size.

Table 4: Reasons for the non-adoption of the 38 management practices (as a % of all practices), before and after treatment

Non-adoption reason	Group	Management practice type	Timing relative to treatment: 1 month before	1 month after	3 months after	5 months after	7 months after	9 months after
Lack of information (plants never heard of the practice before)	Treatment	Common	3.3	3.2	0.5	0	0	0
	Treatment	Un-common	64.0	19.1	2.9	1.5	0	0
	Control	Common	1.9	0	0	0	0	0
	Control	Un-common	67.8	23.7	22.0	22.0	22.0	22.0
Incorrect information (heard of the practice before but think it is not be worth doing)	Treatment	Common	30	22.4	15.4	15.2	14.4	14.4
	Treatment	Un-common	30.9	50.7	50.7	49.3	49.3	47.1
	Control	Common	18.5	18.5	18.5	18.5	18.5	18.5
	Control	Un-common	27.1	52.5	50.9	50.9	49.2	49.2
Owner time, ability or procrastination (the owner is the reason for non-adoption)	Treatment	Common	1.1	0.8	0.5	0.8	1.6	0.8
	Treatment	Un-common	3.7	13.2	13.2	13.2	13.2	14.0
	Control	Common	3.7	3.7	3.7	3.7	3.7	3.7
	Control	Un-common	3.4	20.3	18.6	18.6	18.6	18.6
Other (variety of other reasons)	Treatment	Common	0	0	0	0	0	0
	Treatment	Un-common	2.1	1.5	1.5	2.2	2.2	2.2
	Control	Common	0	0	0	0	0	0
	Control	Un-common	0	0	0	0	0	0
Total non-adoption	Treatment	Common	34.6	26.4	16.3	16.0	16.0	15.2
	Treatment	Un-common	98.5	84.6	78.2	66.2	65.1	63.2
	Control	Common	25.1	22.2	22.2	22.2	22.2	22.2
	Control	Un-common	98.3	96.6	91.5	91.5	89.8	89.8

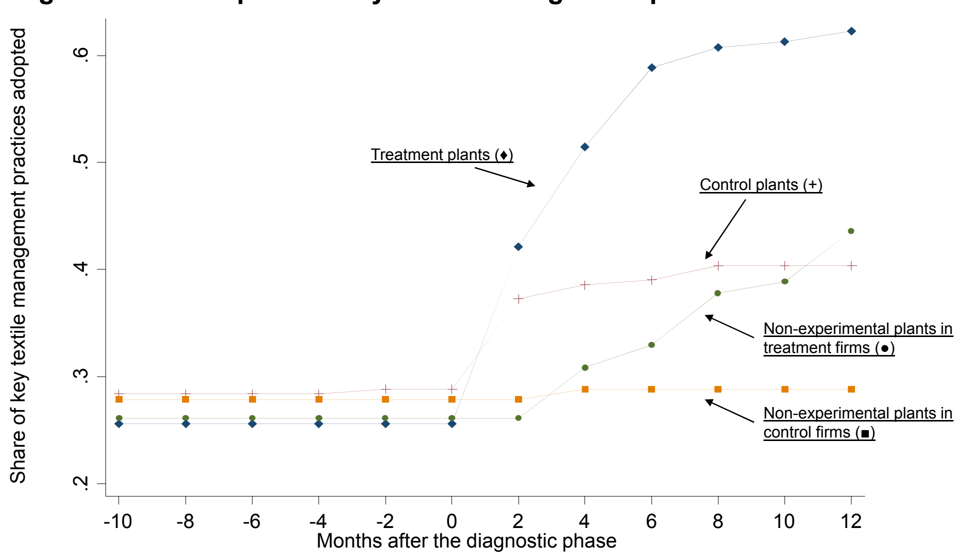
Notes: Percentages (%) of practices not adopted by reason. **Common** practices are the 8 practices with more than 50% initial adoption, mainly quality and downtime recording, and worker bonuses (see table A1 for details). **Un-common** practices are the 10 practices with less than 5% initial adoption, mainly quality, inventory and downtime review meetings and manager incentive schemes. Timing is relative to the start of diagnostic phase. Covers 532 practices in the treatment plants (38 practices in 14 plants), and 228 practices in the control plants (38 practices in 6 plants). Non adoption was monitored every other month using the tool shown in Exhibit 8, based on discussions with the firms' directors, managers, workers, plus regular consulting work in the factories.

Figure 1: Management practice scores across countries



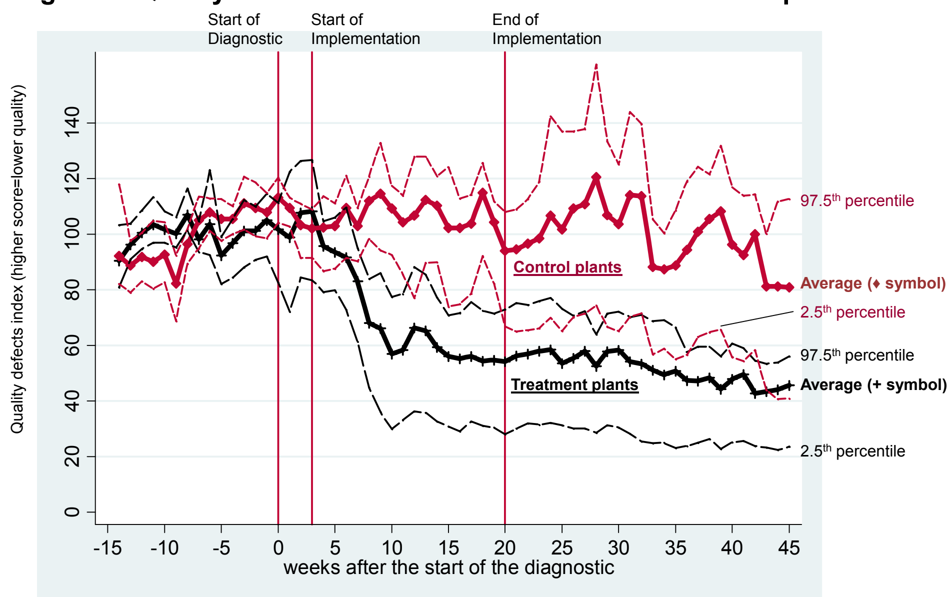
Notes: Management practice histograms using Bloom and Van Reenen (2007) methodology. Double-blind surveys used to evaluate firms' monitoring, targets and operations. Scores from 1 (worst practice) to 5 (best practice). Samples are 695 US firms, 620 Indian firms, 1083 Brazilian and Chinese firms, 232 Indian textile firms and 17 experimental firms.

Figure 2: The adoption of key textile management practices over time



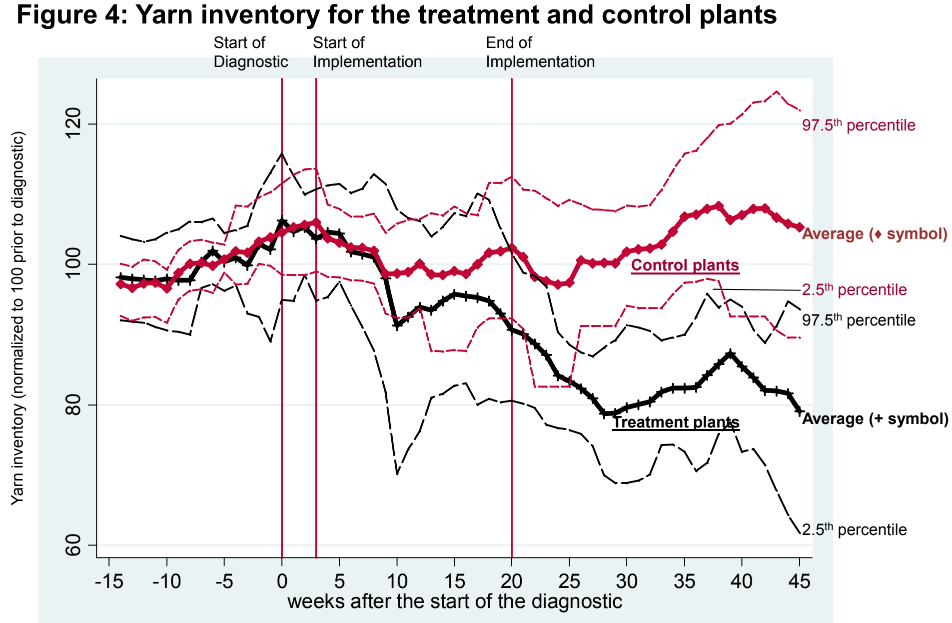
Notes: Average adoption rates of the 38 key textile manufacturing management practices listed in Table 2. Shown separately for the 14 treatment plants (diamond symbol), 6 control plants (plus symbol), the 5 non-experimental plants in the treatment firms which the consultants did not provide any direct consulting assistance to (round symbol) and the 3 non-experimental plants in the control firms (square symbol). Scores range from 0 (if none of the group of plants have adopted any of the 38 management practices) to 1 (if all of the group of plants have adopted all of the 38 management practices). Initial differences across all the groups are not statistically significant.

Figure 3: Quality defects index for the treatment and control plants



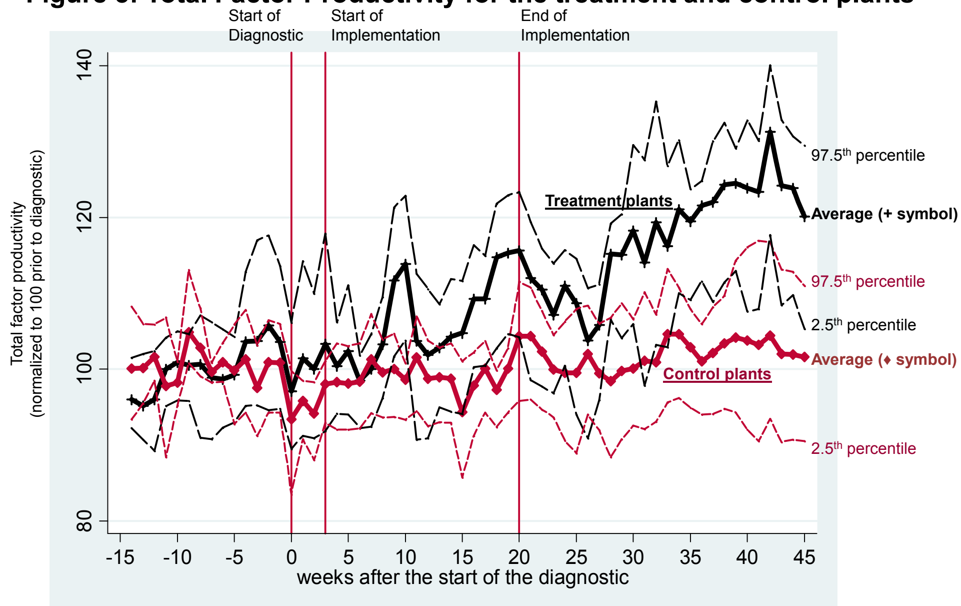
Notes: Displays the average weekly quality defects index, which is a weighted index of quality defects, so a higher score means lower quality. This is plotted for the 14 treatment plants (+ symbols) and the 6 control plants (♦ symbols). Values normalized so both series have an average of 100 prior to the start of the intervention. To obtain confidence intervals we bootstrapped the firms with replacement 250 times. Note that seasonality due to Diwali and the wedding season impacts both groups of plants.

Figure 4: Yarn inventory for the treatment and control plants



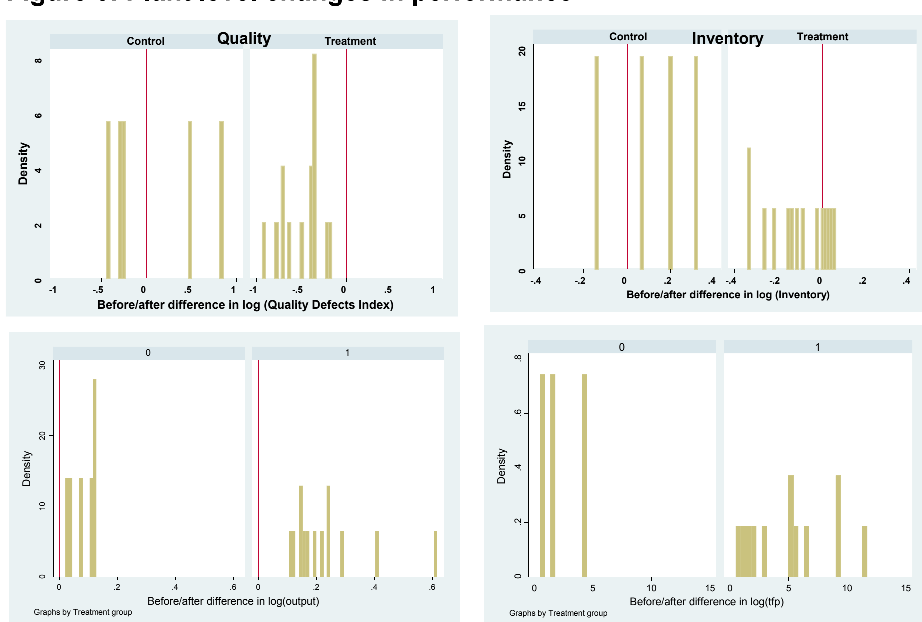
Notes: Displays the weekly average yarn inventory plotted for 12 treatment plants (+ symbols) and the 6 control plants (♦ symbols). Values normalized so both series have an average of 100 prior to the start of the intervention. To obtain confidence intervals we bootstrapped the firms with replacement 250 times. 2 treatment plants maintain no on-site yarn inventory. Note that seasonality due to Diwali and the wedding season impacts both groups of plants.

Figure 5: Total Factor Productivity for the treatment and control plants



Notes: Displays the weekly average TFP for the 14 treatment plants (+ symbols) and the 6 control plants (♦ symbols). Values normalized so both series have an average of 100 prior to the start of the intervention. Confidence intervals we bootstrapped the firms with replacement 250 times. Note that seasonality due to Diwali and the wedding season impacts both groups of plants.

Figure 6: Plant level changes in performance



Notes: Displays the histogram of plant by plant changes in log (Quality Defects Index), log (Inventory) and log (Real Output) and log(TFP) between the post and pre treatment periods.