DOES MANAGEMENT MATTER? EVIDENCE FROM INDIA

Author(s): Nicholas Bloom, Benn Eifert, Aprajit Mahajan, David McKenzie and John Roberts

Source: *The Quarterly Journal of Economics*, Vol. 128, No. 1 (February 2013), pp. 1-51

Published by: Oxford University Press

Stable URL: https://www.jstor.org/stable/10.2307/26372493

THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. 128 February 2013 Issue 1

DOES MANAGEMENT MATTER? EVIDENCE FROM INDIA*

Nicholas Bloom
Benn Eifert
Aprajit Mahajan
David McKenzie
John Roberts

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 17% in the first year through improved quality and efficiency and reduced inventory, and within three years led to the opening of more production plants. Why had the firms not adopted these profitable practices previously? Our results suggest that informational barriers were the primary factor explaining this lack of

*Financial support was provided by the Alfred Sloan Foundation, the Freeman Spogli Institute, the International Initiative, the Graduate School of Business at Stanford, the International Growth Centre, the Institute for Research in the Social Sciences, 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, Rohan Dhote, Karl Gheewalla, Kusha Goyal, Manish Makhija, Abhishek Mandvikar, Shruti Rangarajan, Jitendra Satpute, Shreyan Sarkar, Ashutosh Tyagi, and Ravindra Vasant; and the research support of Troy Smith. We thank the editor, Larry Katz; six anonymous referees; 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 a large number of seminar audiences.

The Quarterly Journal of Economics (2013), 1–51. doi:10.1093/qje/qjs044.
Advance Access publication on November 18, 2012.

adoption. Also, because reallocation across firms appeared to be constrained by limits on managerial time, competition had not forced badly managed firms to exit. *JEL* Codes: L2, M2, O14, O32, O33.

I. INTRODUCTION

Economists have long puzzled over why there are such astounding differences in productivity across both firms and countries. For example, U.S. plants in industries producing homogeneous goods like cement, block ice, and oak flooring display 100% productivity spreads between the 10th and 90th percentile (Foster, Haltiwanger, and Syverson 2008). This productivity dispersion appears even larger in developing countries, with Hsieh and Klenow (2009) estimating that the ratio of the 90th to the 10th percentiles of total factor productivity is 5.0 in Indian and 4.9 in Chinese firms.

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), is emphasized by Leibenstein (1966), and is central to the Lucas (1978) model of firm size. Although management has long been emphasized by the media, business schools, and policy makers, 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, including quality control procedures, inventory management, and certain human resources 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, profitable firms may simply find it easier to adopt better management practices.

This article provides the first experimental evidence on the importance of management practices in large firms. The experiment took large, multiplant Indian textile firms and randomly allocated 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 within the first year productivity increased by 17%; based on these changes we impute that annual profitability increased by over $300,000. These better-managed firms also appeared to grow faster, with suggestive evidence that better management allowed them to delegate more and open more production plants in the three years following the start of the experiment. These firms also spread these management improvements from their treatment plants to other plants they owned, providing revealed preference evidence on their beneficial impact.

Given this large positive impact of modern management, the natural question is why firms had not previously adopted these practices. Our evidence, though speculative, suggests that informational constraints were the most important factor. For many simple, already 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

1. See for example the extensive surveys in Bloom and Van Reenen (2011) and Lazear and Oyer (2012). 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, whereas Locke, Qin, and Brause (2007) show better management practices are associated with improved worker treatment, and Bloom et al. (2010) show better management practices are associated with more energy efficient production.

was as good as that of other (local) firms and, because 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. Although these types of lean management practices are common in Japan and the United States, they appear to be rare in developing countries.

Why did competition not force badly run firms to exit? The reason appears to be that competitive pressures were heavily restricted: imports by high tariffs, entry by the lack of external finance, and reallocation by limited managerial time. Managerial time was constrained by the number of male family members. Non–family members were not trusted by firm owners with any decision-making power, and as a result firms did not expand beyond the size that could be managed by close (almost always male) family members. Not surprisingly, we found that the number of male family members had more than three times the explanatory power for firm size as their management practices.

The major challenge of our experiment was 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 implemented permutation tests whose properties are independent of sample size. We also exploited 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 and also potentially for other field experiments where the data has a small cross-section but long time-series dimension.

This article relates to several strands of literature. First, there is the large body of 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, Syverson's (2011) recent survey of the productivity literature concludes 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 multiplant firms (McKenzie 2010).

Second, our article 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 based on case studies or surveys, making it hard to distinguish between different explanations and resulting in little consensus in the management literature. This article provides experimental evidence suggesting that there is 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 (e.g., Drexler, Fischer, and Schoar 2010; Karlan and Valdivia 2011; Bruhn and Zia 2011; Bruhn, Karlan, and Schoar 2012; Karlan, Knight, and Udry 2012) have begun to estimate the impact of basic business training and advice on micro- and small enterprises.[2] This research has so far delivered mixed results. Some studies find significant effects of business training on firm performance although other studies find no effect. The evidence suggests that differences in the quality and intensity of training, and the size of the recipient enterprises are important factors determining the impact of business training. Our research builds on this literature by providing high-quality management consulting to large, multiplant organizations.

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 United States (Caselli 2011), with a large variation in productivity, spanning a few highly productive firms and many low-productivity firms (Hsieh and Klenow 2009).

In common with other developing countries for which data are available, Indian firms are also typically poorly managed. Evidence of this is seen in Figure I, which plots results from the Bloom and Van Reenen (2010) (henceforth BVR) surveys of

2. See McKenzie and Woodruff (2012) for an overview of business training evaluations.

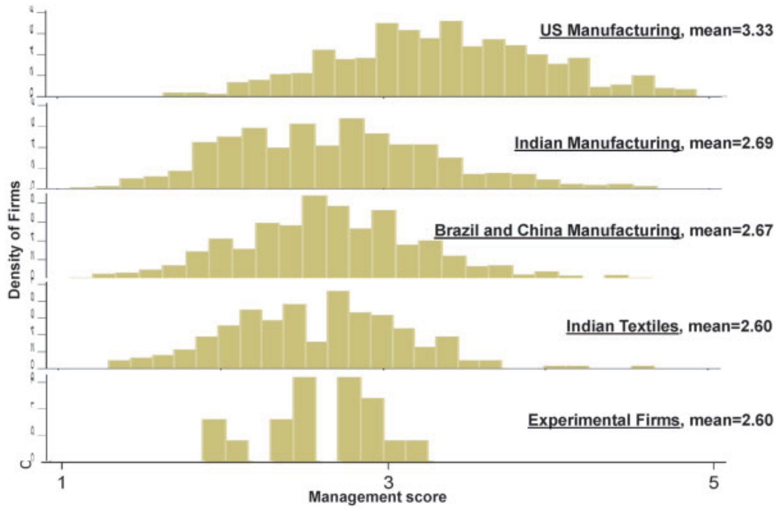

FIGURE I

Management Practice Scores across Countries

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 U.S. firms, 620 Indian firms, 1,083 Brazilian and Chinese firms, 232 Indian textile firms, and 17 experimental firms. Data from http://www.worldmanagementsurvey.com.

manufacturing firms in the United States and India. The 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 I plots these management practice scores for a sample of 695 randomly chosen U.S. manufacturing firms with 100 to 5,000 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 U.S. 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 the management of Indian firms is broadly representative of large firms in emerging economies.

To implement a common set of management practices across firms in our field experiment and measure a common set of outcomes, we focused on one industry. We chose textile production because 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 in 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 I, 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 major developing countries more generally.

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 around Mumbai, based on lists provided by the Ministry of Corporate Affairs (MCA).[3] We restricted attention to firms with between 100 to 1,000 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) because this reduced the travel time for the consultants. This yielded a sample of 66 potential subject firms.

All of these 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

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 manufacturing firms of more than 100 employees, which are too large and permanent to avoid government detection.

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 copayments, 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 these 34 firms, 17 agreed to commit senior management time to the consulting program.[4] We refer to these firms in the subsequent discussion as *project firms*.

This of course generates a selection bias in that our results are valid only for the sample of firms that selected into the experiment (Heckman 1992). We took two steps to assess the extent of the bias. First, we compared the project firms with the 49 nonproject firms and found no significant differences, at least in observables.[5] Second, in late 2011 we ran a detailed ground-based survey of every textile firm around Mumbai with 100 to 1,000 employees (see Online Appendix A2 for details). We identified 172 such firms and managed to interview 113 of them (17 project firms and 96 nonproject firms). The interviews took place at the firms' plants or headquarters and focused on ownership, size, management practices, and organizational data from 2008 to 2011. We found the 17 project firms were not significantly different in terms of preintervention observables from the 96 nonproject firms that responded to this survey.[6]

Although the previous results are comforting in that our treatment and control plants appeared similar to the industry

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 (one day a week). It is also possible these firms were suspicious of the offer, given many firms in India have tax and regulatory irregularities.

5. These observables for project and nonproject firms are total assets, employee numbers, total borrowings, and the BVR management score, with values (p-values of the difference) of \$12.8m versus \$13.9m (.841), 204 versus 221 (.552), \$4.9m versus \$5.5m (.756), and 2.52 versus 2.55 (.859), respectively.

6. These observables for project and nonproject firms included age, largest plant size in 2008 (in loom numbers), largest plant size in 2008 (in employees), and adoption of basic textile management practices in 2008 (see Online Appendix Table AI) with values (p-values of the difference) of 22 versus 22.6 years (.796), 38 versus 42 looms (.512), 93 versus 112 employees (.333), and 0.381 versus 0.324 practice adoption rates (.130), respectively. We compared these values across the 17 project firms and the 96 nonproject firms using 2008 data to avoid any effect of the experiment.

preintervention along observables, there is still the potential issue that selection into the experimental sample was driven by unobservables. We cannot rule out this possibility, though we note that the sign of the bias is ambiguous—the experimental effect may be larger than the effect in the general population if firms with more to gain are more likely to participate, or it may be smaller if firms with the most to gain from improvement are also the most skeptical of what consultants can do.[7] Nevertheless, because typical policy efforts to offer management training to firms will also rely on firms volunteering to participate, we believe our estimate of the effect of improving management is policy relevant for the types of firms that take advantage of help when it is offered.

II.C. The Characteristics of the Experimental Firms

The experimental firms had typically been in operation for 20 years and all were family-owned.[8] They all produced fabric for the domestic market (although some also exported). Table I reports summary statistics for the textile manufacturing parts of these firms (many of the firms have other businesses in textile processing, retail, and even real estate). On average these firms had about 270 employees, assets of $13 million, and sales of $7.5 million a year. Compared to U.S. manufacturing firms, these firms would be in the top 2% by employment and the top 4% by sales, and compared to India manufacturing they are in the top 1% by both employment and sales (Hsieh and Klenow 2010). Hence, these are large manufacturing firms by most standards.[9]

These firms are also complex organizations, with a median of two plants per firm (plus a head office in Mumbai) and four reporting levels from the shop floor to the managing director. In all the firms, the managing director was the largest shareholder, and all directors were family members. Two firms were publicly quoted on the Mumbai Stock Exchange, although

7. There is now some evidence on the importance of self-selection in laboratory experiments. Harrison, Lau, and Rustrom (2009) find that these effects are relatively small in the class of experiments they examined, whereas Lazear, Malmendier, and Weber (2012) find stronger evidence of self-selection into experiments on social preferences.

8. Interestingly, every single firm in our 113 industry sample was also family-owned and managed.

9. Note that most international agencies define *large firms* as those with more than 250 employees.

TABLE I
THE FIELD EXPERIMENT SAMPLE

	All				Treatment	Control	Diff
	Mean	Median	Min	Max	Mean	Mean	p-value
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
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	8.50	5.21	1.89	29.33	8.83	7.96	0.837
Daily mtrs, experimental plants	5,560	5,130	2,260	13,000	5,757	5,091	0.602
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
Quality defects index	5.24	3.89	0.61	16.4	4.47	7.02	0.395
Inventory (1,000 kilograms)	61.1	72.8	7.4	117.0	61.4	60.2	0.945
Output (picks, million)	23.3	25.4	6.9	32.1	22.1	25.8	0.271
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 nonexperimental 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 textile 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, experimental plants* 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 five hierarchical levels. *Annual sales ($m)* and *Current assets ($m)* are both in 2009 US$ million values, exchanged at 50 rupees = US$1. *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 1,600 suits daily. *BVR management score* is the Bloom and Van Reenen (2007) management score for the experimental plants. *Management adoption rates* are the adoption rates of the management practices listed in Appendix Table A.I in the experimental plants. *Age of experimental plant (years)* reports the age of the plant for the experimental plants. *Quality defect index* is a severity weighted measure of production quality defects. *Inventory* is the stock of yarn per intervention. *Output* is the production of fabric in picks (one pick is a 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 are pooled across pre–diagnostic phase data.

more than 50% of the equity in each was held by the managing family.

In Figures II, III, and IV and Exhibits O1 to O4 in the Online Appendix, we include a set of photographs of the plants. These are included to provide some background information on their size, production processes, and initial state of management. Each plant site involved several multistory buildings and operated continuously—24 hours a day (in two 12-hour shifts), 365 days a

FIGURE II

Many Parts of These Factories Were Dirty and Unsafe

FIGURE III

The Factory Floors Were Frequently Disorganized

FIGURE IV

Most Plants Had Months of Excess Yarn, Usually Spread across Multiple Locations, Often without Any Rigorous Storage System

year. The factory floors were often dirty (Figure II) and disorganized (Figure III), and their yarn and spare parts inventory stores frequently lacked any formalized storage systems (Figure IV). This disorganized production led to frequent quality defects (oil stains, broken threads, wrong colors, etc.) necessitating an extensive checking and mending process that employed 19% of the factory manpower, on average.

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) and measure the impact of doing so on firm performance. To achieve this, we hired a management consultancy firm to work with the plants as the easiest way to change plant-level management rapidly. We selected the consulting firm using an open tender. The winner was a large international management consultancy that is headquartered in the United States and has about 40,000 employees in India. The full-time team of (up to) six 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 U.S. and European multinationals.

Selecting a high-quality international consulting firm substantially increased the cost of the project.[10] 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 first (and main) wave of the project ran from August 2008 to August 2010, with a total consulting cost of $1.3 million, approximately $75,000 per treatment plant and $20,000 per control plant. This is different from what the firms themselves would have to pay for this consulting, which the consultants indicated would be about $250,000. The reasons for our lower costs per plant are that the consultancy firm 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. The second wave ran from August 2011 to November 2011, with a total consulting cost of $0.4 million, and focused on collecting longer-run performance and management data. The intention to undertake this wave was not mentioned in August 2010 (when the first wave finished) to avoid anticipation effects.

Although 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.

10. At the bottom of the consulting quality distribution in India consultants are cheap, but their quality is poor. At the top end, rates are similar to those in the United States because international consulting companies target multinationals and employ consultants who are often U.S.- or European-educated and have access to international labor markets.

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 U.S., European, and Japanese firms, and can be grouped into following 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 facilitate the movement of materials. Establishing standard procedures for operations.
- 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 resources 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 practices (listed in Appendix Table A.I) 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.[11] 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 routinely use them in

11. We prefer these indicators to the BVR management score for our work here, because they are all binary indicators of specific practices that are directly linked to the intervention. In contrast, the BVR indicator measures practices at a more general level on a five-point ordinal scale. Nonetheless, the sum of our 38 preintervention management practice scores is correlated with the BVR score at 0.404 (p-value of .077) across the 17 project firms.

decisions. Thus, although 93% of the treatment plants recorded quality defects before the intervention, only 29% monitored them on a daily basis or by the particular sort of defect, and none had any standardized system to analyze and act on this data.

The consulting treatment had three phases. The first phase, 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, called 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 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 phase 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 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 employees could readily carry them out. For example, one of the practices was holding 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 in the first wave until August 2010, and then in the second (follow-up) wave from August 2011 to November 2011. This involved collection of performance and management data from all treatment and control plants. In return for this continuing data, the consultants provided light consulting advice to the treatment and control plants. This phase involved about 1.5 days a month of consulting per plant.

In summary, the control plants were provided with the diagnostic phase and then the measurement phases (totaling 273 consultant hours on average), and the treatment plants were provided with the diagnostic, implementation, and then measurement phases (totaling 781 consultant hours on average).

III.C. The Experimental Design

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

1. 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. Two factors pushed against this. First, many large firms in India are reluctant to let outsiders into their plants, probably because of compliance issues with various regulations. To minimize selection bias, we offered a high-quality intensive consulting intervention that firms would value enough to risk 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 sample size. We discuss the estimation issues this generates in Section III.D.

2. Treatment and control plants. The 17 firms that agreed to participate in the project had 28 plants between them. Of these, 25 were eligible to be treatment or control plants because we could obtain historic performance data from them. Randomization occurred at the firm level and was conducted by computer. We first randomly chose six firms to be the control firms, and one eligible plant from each of them to be the control plants. The remaining 11 firms were then the treatment firms. Our initial funding and capacity constraints of the consulting team meant that we could start with four plants as a first round, which started in September 2008. We therefore randomly chose 4 of the 11 treatment firms to be in round 1, randomly selecting one plant from each firm. In April 2009, we started a second round of treatment. This comprised selecting a random plant from each of the remaining seven treatment firms, and, because funding allowed for it, three more plants selected at random from the treatment firms with multiple plants. Pure randomization, rather than stratification or rerandomizing, was used in each step. This was done both because of initial uncertainty as to how many plants we would have funding to treat and because of concerns about variance estimation and power when stratified randomization is used in very small samples (Bruhn and McKenzie 2009).

The result is that we have 11 treatment firms, with 14 treatment plants among them, and 6 control firms, each with a control plant. 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 other and because we believed the treatment plants would be more useful for understanding why firms had not adopted management practices before. Table I shows that the treatment and control firms were not statistically different across any of the characteristics we could observe.[12] The remaining eight plants were then classified as "nonexperimental plants": three in control firms and five in treatment firms. These nonexperimental plants did not directly receive any consulting services, but data on their management practices were collected in bimonthly visits.

3. Timing. The consulting intervention was executed in three rounds because of the capacity constraint of the six-person consulting team. The first round started in September 2008 with four treatment plants. In April 2009 a second round with 10 treatment plants was initiated, and in July 2009 the final round with 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 round because we expected the intervention process to get easier over time due to accumulated experience. The second round included all the remaining

12. We test for differences in means across treatment and control plants, clustering at the firm level.

treatment firms because (1) the consulting interventions take time to affect performance and we wanted the longest time window to observe the treatment firms, and (2) we could not mix the treatment and control firms across implementation rounds.[13] The third round contained the control firms.

III.D. Small Sample Size

The focus on large firms meant we had to work with a small number 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. The third potential concern is whether the sample is 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.

1. Significance of results. Even though we had only 20 experimental plants across the 17 project firms, we obtained 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—tended to average out. Second, the data were collected directly from the machine logs, and had very little (if any) measurement error. Third, the firms were homogeneous in terms of size, product, region, and technology, so time dummies controlled for most external shocks. Fourth, we collected weekly data, which provided high-frequency observations over the course of the treatment. The use of these repeated measures can reduce the sample size needed to detect a given treatment effect, although this is tempered by the degree of serial correlation in the output data (McKenzie 2012), which was around 0.8 for our performance metrics. Finally, the intervention was intensive, leading to large treatment effects—for example, the point estimate for the reduction in quality defects was almost 50%.

13. Each round 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. Because this meeting involved a project outline, and we did not tell firms about the different treatment and control programs, we could not mix the groups in the meetings.

2. Statistical inference. A second concern is over using statistical tests that rely on asymptotic arguments in the cross-sectional dimension (here, the number of firms) to justify the normal approximation. We use three alternatives to address this concern. First, we use firm-clustered bootstrap standard errors (Cameron, Gelbach, and Miller 2008). Second, we implement permutation procedures (for both the intention to treat and instrumental variables estimators) that do not rely on asymptotic approximations. Third, we exploit our large T sample to implement procedures that rely on asymptotic approximations along the time dimension (with a fixed cross-sectional dimension).[14]

3. 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 focused 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 5,000 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 I). In Online Appendix Figure OI, we also plot results on a plant-by-plant basis to further demonstrate that the results are not driven by any particular plant outlier. Although we have a small sample, the results are relatively stable across the individual sample plants.

III.E. Potential Conflict of Interest

A final design challenge was the potential for a conflict of interest in having our consulting firm measuring the performance of the experimental plants. To address this, we first 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 member of the research team visited the firms, met with the directors, and presented the quality, inventory, and output data the consultants had sent us. This was positioned as a way to initiate discussions on

14. These permutation and large T procedures are summarized in the Appendix and detailed in Online Appendix B.

the impact of the experiment with the directors, but it also served to check that the data we were receiving reflected reality. We would likely have received some pushback if the results had been at variance with the owners' own judgment. Finally, some of the long-run data, like the number of plants, is directly observable, so it would be hard for the consulting firm to fabricate this.

IV. THE IMPACT ON MANAGEMENT PRACTICES

In Figure V, we plot the average management practice adoption of the 38 practices for the 14 treatment plants, the 6 control plants, the 8 nonexperimental plants, and our 96 nonproject firms surveyed in 2011. This management practice score is the proportion of the 38 practices a plant had adopted. This data for the project firms is shown at 2-month intervals starting 10 months before the diagnostic phase and extending to at least 24 months after. The nonproject firm data was collected at a yearly frequency using retrospective information. For the project firms, data from the diagnostic phase onward were compiled from direct observation at the factory, and data from before the diagnostic phase were collected from detailed retrospective interviews of the plant management team. For the nonproject firms, data were collected during the interview from direct factory observation and detailed discussion with the managers (details in Online Appendix AII). Figure V shows six results.

First, all plants started off with low baseline adoption rates of the set of management practices.[15] Among the 28 individual plants in the project firms, 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 had 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 I.[16] 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

15. The pretreatment difference between the treatment, control, and other plant groups is not statistically significant, with a p-value on the difference of .550 (see Appendix Table A.I).

16. Interestingly, Clark (1987) suggests Indian textile plants may have even been badly managed in the early 1900s.

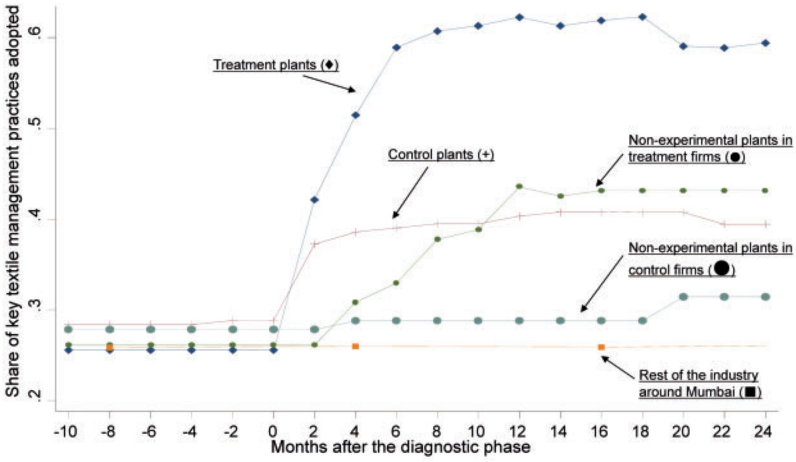

Figure V

The Adoption of Key Textile Management Practices over Time

Average adoption rates of the 38 key textile manufacturing management practices listed in Table A.I. Shown for the 14 treatment plants (diamond), 6 control plants (plus sign), the 5 nonexperimental plants in the treatment firms to which the consultants did not provide any direct consulting assistance (small circle), the 3 nonexperimental plants in the control firms (large circle), and 96 plants from the rest of the industry around Mumbai (square). 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. The 96 plants from the rest of the industry were given the same diagnostic phase start date as the control plants (July 2009).

repeatedly. Most also had not organized their yarn inventories, so yarn stores were frequently 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 by 37.8 percentage points on average by August 2010, when the main wave ended (an increase from 25.6% to 63.4%). These improvements in management practices were also persistent. The management practice adoption rates dropped by only 3 percentage points, on average, between the end of the first wave in August 2010 (when the consultants left) and the start of the second wave in August 2011.

Third, not all practices were adopted. The firms appeared to adopt the practices that were the easiest to implement and/or had the largest perceived 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.[17]

Fourth, the treatment plants' adoption of management practices occurred gradually and nonuniformly. In large part, this reflects the time taken for the consulting firm to gain the confidence of the 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 human resources.

Fifth, the control plants, which were given only the one-month diagnostic, increased their adoption of the management practices, but by only 12 percentage points, 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. However, it is an increase relative to the rest of the industry around Mumbai (the nonproject plants), which did not change their management practices on average between 2008 and 2011.

Finally, the nonexperimental plants in the treatment firms also saw a substantial increase in the adoption of management practices. In these five plants, the adoption rates increased by 17.5 percentage points by August 2010. This increase occurred because the directors of the treatment firms copied the new practices from their experimental plants over to their other plants. Interestingly, this increase in adoption rates is similar to the control firms' 12 percentage point increase, suggesting that copying best practices across plants within firms can be as least as effective at improving management practices as short (one-month) bursts of external consulting.

17. See, for example, Suri (2011) for a related finding on heterogeneous agricultural technology adoption in Kenya.

V. THE IMPACT OF MANAGEMENT ON PERFORMANCE

V.A. Intention to Treat Estimates

We estimate the impact of the consulting services on management practices via the following intention to treat (ITT) equation:

$$OUTCOME_{i,t} = aTREAT_{i,t} + bDURING_{i,t} + c_t + d_i + e_{\mathrm{i},t}, \tag{1}$$

where *OUTCOME* is one of the key performance metrics of quality, inventory, output, and total factor productivity (TFP).[18] 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 (2004–5), capital includes all physical capital (land, buildings, equipment, and inventory), and labor is production hours (see Online Appendix AI for details). $TREAT_{i,t}$ takes the value of one for the treatment plants starting one month after the end of the intervention period and until the end of the study and is zero otherwise, while $DURING_{i,t}$ takes the value of one for the treatment plants for the six-month window from the start of the diagnostic period. The c_t are a full set of weekly time dummies to control for seasonality, and the d_i are a full set of plant dummies that were included to control for 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 gives the ITT, which is the average impact of the implementation in the treated plants, and b shows the short-term impact during the implementation.[19]

In addition to this specification, in Appendix Table A.II we estimate the impact on outcomes of our index of management practices, using the consulting services as an instrument, and

18. We study quality, inventory, and output because these are relatively easy to measure key production metrics for manufacturing. They also directly influence TFP because 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.

19. In the case that a varies across plants, our estimate of a will be a consistent estimate of the average value of a_i. Note that because the diagnostic was received by both treatment and control plants, the ITT estimates the effect of the implementation on treatment plants in a situation where both treatment and control plants receive the diagnostic.

compare the instrumental variables (IV) results to the fixed-effects estimates. We find that the fixed-effects estimates tend to understate the gain in performance from better management, which is consistent with changes in management being more likely to be implemented when outcomes are declining.

We use performance data up to the start of September 2010, since the data are not comparable after this date because of investment in new looms in some treatment plants. The firm directors began replacing older Sulzer and Rapier looms with Jacquard looms, which produce higher mark-up fabric but require more advanced quality control and maintenance practices. This started in September 2010 after the end-of-summer production surge for the wedding season and the Diwali holiday.

In Table II column (1) we see that the ITT estimate for quality defects shows a significant drop of 25% occurring just during the implementation period, eventually falling further to a 43% drop.[20] This is shown over time in Figure VI, which plots the QDI score for the treatment and control plants relative to the start of the treatment period: September 2008 for round 1 treatment and April 2009 for round 2 treatment and control plants.[21] The score is normalized to 100 for both groups of plants using pretreatment data. To generate point-wise confidence intervals we block-bootstrapped over the firms.

The treatment plants started to reduce their QDI scores (i.e., improve quality) significantly and rapidly from about week 5 onward, which was the beginning of the implementation phase following the initial one-month diagnostic phase. The control firms also 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 likely reason for this huge reduction in defects is that measuring, classifying, and tracking 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

20. Note that quality is estimated in logs, so that the percentage reduction is $-43.1 = \exp(-0.564) - 1$.

21. Because the control plants have no treatment period, we set their timing to zero to coincide with the 10 round 2 treatment plants. This maximizes the overlap of the data.

TABLE II

THE IMPACT OF MODERN MANAGEMENT PRACTICES ON PLANT PERFORMANCE

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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
Intervention$_{i,t}$	−0.564**	−0.245**	0.090**	0.154*				
	(0.235)	(0.117)	(0.037)	(0.084)				
During implementation$_{i,t}$	−0.293**	−0.070	0.015	0.048				
	(0.137)	(0.093)	(0.031)	(0.056)				
Cumulative treatment$_{i,t}$					−0.032**	−0.015**	0.006***	0.009**
					(0.013)	(0.005)	(0.002)	(0.004)
Small sample robustness								
Ibragimov-Mueller (95% CI)	[−1.65,0.44]	[−0.83,-0.02]	[0.05,0.38]	[−0.014,0.79]				
Permutation test (p-value)	.001	.060	.026	.061				
Time FEs	127	127	127	127	127	127	127	127
Plant FEs	20	18	20	18	20	18	20	18
Observations	1,807	2,052	2,393	1,831	1,807	2,052	2,393	1,831

Notes. All regressions use a full set of plant and calendar week dummies. Standard errors are bootstrap clustered at the firm level. *Intervention* is a plant level dummy equal to 1 after the implementation phase at treatment plants and 0 otherwise. *During implementation* is a dummy variable equal to 1 from the beginning of the diagnostic phase to the end of the implementation phase for all treatment plants. *Cumulative treatment* is the cumulative weeks of treatment since the beginning of the implementation phase in each plant (0 in both the control group and prior to the implementation phase in the treatment group). *Quality defects* is the 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. *ITT* reports the intention to treat results from regressing the dependent variable directly on the intervention dummy. *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 are available. *Small sample robustness* implements two different procedures (described in greater detail in the Appendix and Online Appendix B) to address issues of plant heterogeneity, within plant (and firm) correlation, and small sample concerns. *Ibragimov-Mueller (95% CI)* report 95% confidence interval estimates from firm-by-firm parameter estimates treating the estimates as draws from independent (but not identically distributed) normal distributions and conducts a two-sample t-test. *Permutation test* 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 12,376 possible permutations of treatment assignment. These tests have exact finite sample size. *** denotes 1%, ** denotes 5%, * denotes 10% significance.

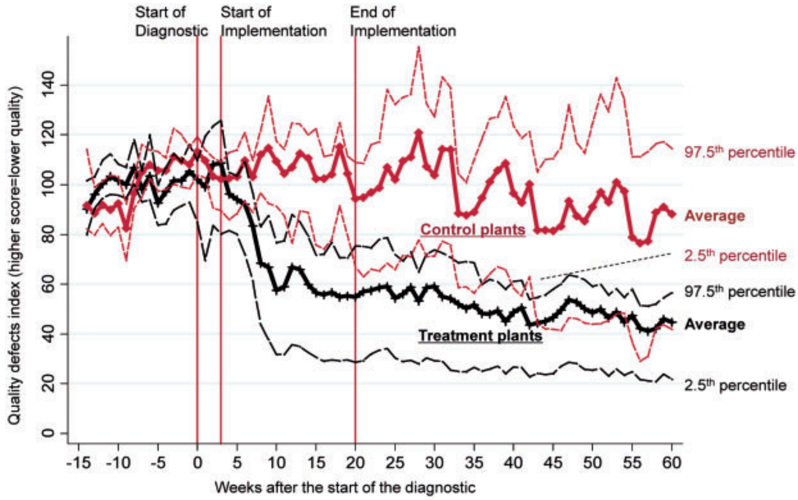

FIGURE VI

Quality Defects Index for the Treatment and Control Plants

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 (plus signs) and the 6 control plants (diamonds). 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 affects both groups of plants.

and mending team had no mechanism (or incentive) to reduce defects. In the longer term, the QDI also allowed 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 could be dropped from the sales catalog. 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 U.S. automotive plants saw reductions in defects of over 90% following the adoption of lean production systems (see, for example, Womack, Jones, and Roos 1990).

At the bottom of Table II we also present results from our robustness checks: the Ibramigov-Mueller (IM) and permutation tests. The results are consistent with a reduction in quality defects. First, looking at the permutation tests that have exact size, we see that the ITT is significant at the 5% level (the p-value is .01). The IM approach that exploits asymptotics in T rather than N finds that the ITT results are consistent with large improvements in quality though the confidence intervals are wide.

Column (2) reports the results for inventory with a 21.7% (= exp(–0.245) – 1) post-treatment reduction, and no significant change during the implementation phase. Figure VII shows the plot of inventory over time. The fall in inventory in treatment plants occurred because they were carrying about four 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 cataloging, the firms sold or used up the surplus yarn (by incorporation into new designs), and then introduced restocking norms for future purchases and monitored against these weekly. This process took some time, and hence inventories do not respond as quickly as quality to the intervention.

In column (3) of Table II, we look at output and see a 9.4% (= exp(+0.09) – 1) increase in output from the intervention, with no significant change during the implementation period. 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 considerably (we estimate by about 50%). Second, undertaking routine maintenance of the looms and collecting and monitoring breakdown data presumably also helped reduced machine downtime. Visual displays around the factory floor together with the incentive schemes also encouraged 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), we show the results for log TFP reporting a 16.6% (exp(0.154) – 1) increase in the treatment firms posttreatment compared to the control firms. 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 VIII shows the time profile of productivity, showing it took time to see impacts, which is why there is no significant effect during implementation.

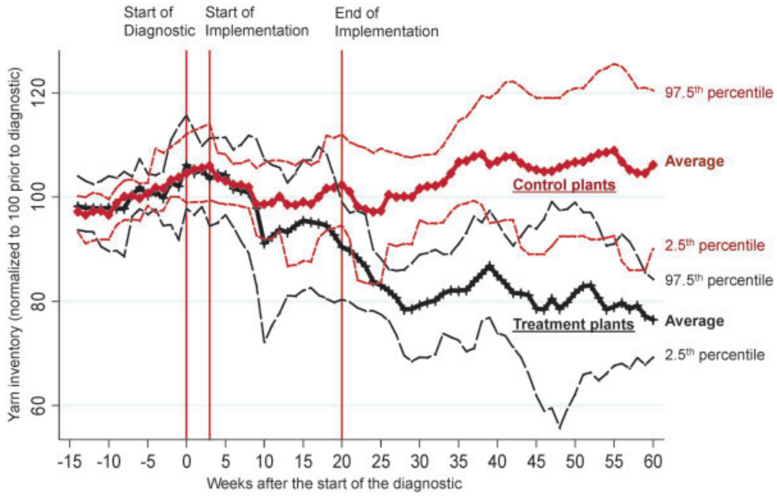

FIGURE VII

Yarn Inventory for the Treatment and Control Plants

Displays the weekly average yarn inventory plotted for 12 treatment plants (plus signs) and the 6 control plants (diamonds). 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. Two treatment plants maintain no on-site yarn inventory. Note that seasonality due to Diwali and the wedding season affects both groups of plants.

In columns (5) to (8) of Table II, 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 V to VIII highlight.

We also plotted the difference in quality, inventory, and output after treatment on a plant-by-plant basis (Online Appendix Figure OI). Although there is some evidence that the results vary across plants, we found no outliers driving these coefficient differences.[22] As another check, we estimated the results with time dummies for weeks relative to the start of the diagnostic phase (rather than calendar time) and found slightly larger

22. Note that the IM procedure also allows for firm-level heterogeneity. In particular, the procedure estimates firm-by-firm coefficients, which form the basis for inference.

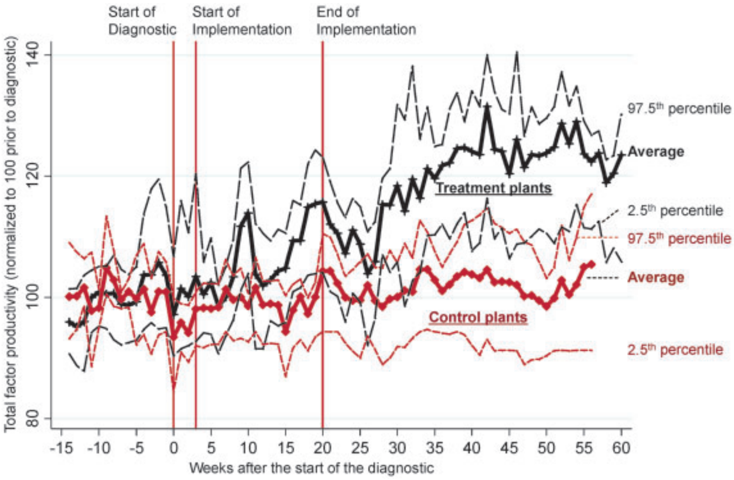

FIGURE VIII

Total Factor Productivity for the Treatment and Control Plants

Displays the weekly average TFP for the 14 treatment plants (plus signs) and the 6 control plants (diamonds). 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 affects both groups of plants.

effects (see Online Appendix Table OI). Finally, we plotted the main performance metrics for each wave by calendar time, again showing gradually rising effects of the intervention on improving quality, reducing inventory, and raising TFP (see Online Appendix Figures OII to OIV).

Using the results from Table II, we estimate a total increase in profits of around $325,000 per plant per year (detailed in Online Appendix B). We could not obtain accounting data on these firms' profits and losses. Public accounts data are available only with a lag of two to three years at the firm level (rather than plant, which is what we would want) and may not be completely reliable (according to our interviews). Firms were also reluctant to provide internal accounts, though they indicated that profits were often in the range of $0.5m to $2m per year.[23] So we infer the

23. It is not even clear if firms actually keep correct records of their profits, given the risks they entail. For example, any employee that discovered such records could use them to blackmail the owners.

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 longer run, the impact might be greater if other complementary changes happen (like firms upgrading 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.[24] So given the $250,000 that the consultancy reported it would have charged an individual firm for comparable services if it paid directly, this implies about a 130% one-year rate of return.

A further question which arises when viewing Figures V through VIII is whether the control firms benefited from the practices they introduced as a result of the diagnostic phase, and if not, why? We first note that we have no counterfactual for the control group, so are unable to say what would have happened if they had not implemented the few practices that they did implement. The fact that we see graphically at most a small gradual decline in quality defect rates for the control group in Figure VI may thus represent an improvement relative to what would have happened absent these new management practices, or else show little return from the specific practices the control firms did implement. Second, Appendix Table A.I shows that the biggest changes in practices for the control firms were largely the results of our measurement—recording quality defect-wise, monitoring machine downtime daily, and monitoring closing stock weekly. Implementing these practices to monitor and electronically record data without implementing the complementary practices to use and act on these data is intuitively likely to have more limited effects. However, because we do not have experimental variation in which practices were implemented, we cannot investigate such complementarities formally.

V.B. Long-run Effects of the Management Intervention

To evaluate the long-run effects of the management intervention, we collected data on the evolution of the number of plants for each firm. The number of plants is a good long-run performance

24. About $35 of extra labor to help organize the stock rooms and factory floor, $200 on plastic display boards, $200 for extra yarn racking, $1,000 on rewards, and $1,000 for computer equipment.

indicator as it is (1) easy to measure and recollect so could be accurately collected over time for the entire set of firms in our 2011 survey, and (2) is not influenced by changes in loom technologies that made comparing plant-level output over time difficult. We also collected data on two other size measures: the number of looms per plant and the number of employees per plant.

In Table III, we see in column (1) that in the 2011 cross-section the number of plants per firm is higher for better managed firms and for firms with more male adult family members. The management variable is the share of the 16 management practices measured in the 2011 survey, which the firms adopted each year.[25] The number of adult male family members is the response to question "how many family members could currently work as directors in the firm?," which aims to record the supply of family managers by including family members currently working the firm plus in any other firms (but excluding those in full-time education or in retirement).[26] In terms of magnitudes, although both indicators are at least weakly significant, we found that the number of family members was much more important in explaining firm size than the management score. The reduction in R-squared from dropping the number of family members was 10.1% (15.9% versus 5.8%), more than three times the reduction of 3.3% (15.9% versus 12.6%) from dropping the management term. This reflects the fact that the number of male family members appeared to be the dominant factor determining firm size, and management practices appeared to play only a secondary role.

In column (2) we regress the time series of firm size in terms of the number of plants in our 17 project firms on a (0/1) post-treatment indicator, including a set of time and firm fixed effects, using yearly data from 2008 to 2011 and clustering by firm. We find a positive coefficient, suggesting that improving management practices in these firms helped them expand, although this is only significant at the 10% level. Treatment firms told us that having better management practices enabled each family

25. These 16 practices are a subset of the 38 practices listed in Appendix Table A.I and were chosen because they could be most accurately measured during a single visit to a plant.

26. We refer to this as male family members because it was extremely rare to have female directors. We came across only one female director in the 113 surveyed firms. She was only running the firm because of her husband's heart attack.

TABLE III

LONG-RUN IMPACT OF THE EXPERIMENT ON FIRM SIZE AND DECENTRALIZATION

	Firm size			Delegation to plant management		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	No. of plants	No. of plants	No. of plants	z-score	z-score	z-score
Sample	Industry	Experiment	Industry	Industry	Experiment	Industry
Time period	2011	2008–2011	2008–2011	2011	2008–2011	2008–2011
Management$_{i,t}$	1.040*			0.597†		
	(0.563)			(0.370)		
Male family members$_{i,t}$	0.210***			0.010		
	(0.065)			(0.042)		
Posttreatment$_{i,t}$		0.217*	0.259**		0.103**	0.171***
		(0.122)	(0.110)		(0.049)	(0.035)
Plant manager related$_i$				0.423***		
				(0.150)		
Plant manager tenure$_i$				0.014**		
				(0.007)		
Small sample robustness						
Permutation tests (p-value)	n/a	0.21	0.02	n/a	0.12	0.001
Time FEs	n/a	3	3	n/a	3	3
Plant/Firm FEs	n/a	17	121	n/a	28	128
Observations	107	68	468	120	108	499

Notes. The *size* dependent variable in columns (1)–(3) is the number of plants in the firm. The *decentralization* dependent variable in columns (4)–(6) is the z-score index of plant decentralization, which is the sum of the four z-scored (normalized to a mean of 0 and standard deviation of 1) individual responses over plant manager autonomy over weaver hiring, junior manager hiring, spare parts purchasing authority, and days the director does not visit the factory (see Online Appendix A.I for details). Columns (1)–(3) are run at the firm level (because firm-size is a firm-level variable) and columns (4)–(6) are run at the plant level (because decentralization is a plant-level variable). *Management* is the adoption share of the 16 management practices starred in Appendix Table A.I and discussed in Online Appendix A.I, averaged across all plants within the same firm in columns (1)–(3). *Male family members* is the number of adult sons and brothers of the interviewed director, which includes all male family members currently working (even working in another firm) but excludes those in school of university. This is designed to measure the supply of male family members that could work in the firm. *Post treatment* takes the value 1 for a treatment firm/plant after the implementation phase and 0 otherwise. *Plant manager related* reports if the plant manager is related to the director, including cousins, uncles, and other indirect family members. *Plant manager tenure* measures the number of years the plant manager has been working at the firm. *Time FEs* report the number of calendar week time fixed effects. *Firm / Plant FEs* reports the number of firm-level fixed effects (columns (1)–(3)) or plant-level fixed effects (columns (4)–(6)). Standard errors clustered at the firm level in all columns. *Permutation test* 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 12,376 possible permutations of treatment assignment. *** denotes 1%, ** denotes 5%, * denotes 10%, †denotes 15% significance.

director to oversee more production activity, as the additional data allowed them to delegate more decisions to their plant managers while closely monitoring them. Expansion occurred through increasing the number of plants (as opposed to expanding current plants) as this apparently reduced the risk of unionization and regulatory problems.[27] Interestingly, we found no impact of management practices on the number of employees per plant or the number of looms per plant.[28] This is because although output was rising at each plant, employment was usually falling (due to a reduction in the amount of mending labor) and loom numbers could move in either direction depending on whether the plant upgraded to Jacquard looms.

One possible reason for the weak significance is the small sample size. In column (3) of Table III we find that our treatment firms are increasing plant numbers in comparison to the whole industry as well as just the control firms. We find a coefficient of 0.259 plants on the treatment dummy, suggesting that over the 3+ years following the start of the intervention our treatment firms opened 0.259 more plants on average than the industry (whose average number of plant openings was 0.120).

In column (4) we investigate the reason for this possible increase in number of plants per firm in terms of the degree of delegation from the directors to managers in the plants. This is measured as the principal factor component of four questions—two on the degree of delegation to plant managers over hiring weavers and hiring managers, one on the rupee investment spending limit of a plant manager, and one on the number of days per week the director visited the plant. We find in column (4), looking first at the cross-section, that more decentralized plants tend to be better managed and to have a plant manager related to the director (i.e., a cousin or uncle) and a long-tenured plant manager. In column (5) we regress delegation on posttreatment and find a significant positive coefficient, suggesting better

27. The owners told us that larger plants attract more regulatory and union attention. For example, both the Factories Act (1947) and Disputes Acts (1947) regulate plants rather than firms, and unionization drives tend to be plant- rather than firm-based.

28. In an identical specification to column (2) in Table III, except with employment and loom number as the dependent variable, we find point estimates (standard errors) of –1.28 (6.19) and 2.38 (3.24), respectively. Interestingly Bruhn, Karlan, and Schoar (2012) also find no effect of consulting on employment despite finding large positive effects on sales and profits.

management leads to more delegation. Finally, in column (6) we compare the change in treatment firms to the whole industry and also find a significant increase in delegation when compared to this larger sample. This is consistent with the story we heard from the directors and the consultants that better management practices enabled the directors to decentralize more decision making to the plant managers, increasing firm size by relaxing the constraint on male family members' management time.

V.C. Are the Improvements in Performance due to Hawthorne Effects?

Hawthorne effects involve the possibility that just running experiments and collecting data can improve performance, raising concerns that our results could be spurious. However, we think these are unlikely to be a major factor in our study, 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 implementation stage while the control plants had briefer but nevertheless frequent visits from the consultants collecting data. Neither the treatment nor control plants were told that two groups existed, so they were not aware of being "treatment" or "control" plants.[29] Hence, it cannot be simply the presence of the consultants or the measurement of performance that generated the relative improvement in performance of treatment firms. 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, and in the longer run led to treatment firms opening additional plants. Third, these improvements persisted after the implementation period,[30] including the long-run change in the number of

29. We told the firms the World Bank wanted to investigate the effect of management practices on performance, so we were providing management advice to the firms and collecting performance data. Thus, the firms believed they were involved in a "difference" rather than a "difference-of-differences" experiment.

30. Note that since we did not inform firms in August 2010 that we would revisit them in August 2011, persistence of management practices due to anticipation effects is unlikely.

plants, so are not some temporary phenomena due to increased attention. Finally, the firms themselves also believed these improvements arose from better management practices, which was the motivation for them extensively copying these practices over to their other nonexperimental plants (see Figure V) and opening new plants.

A related but harder issue is whether the improvements are the lasting impact of temporary increases in human capital. For example, the highly educated consultants could have simply inspired the employees to work harder. Our view is that this is possible, although it was not the main channel of effects from consulting. First, anecdotally, we heard no evidence for this happening from the directors, managers, or consultants. Second, the fact that almost 50% of the management changes were copied across plants suggests that at least part of this was not localized plant-level human capital effects. Third, the fact that the treatment led to the opening of new plants suggests the effects were seen as sufficiently persistent by the directors to justify making long-run investments in new plants, which seems inconsistent with motivation effects which seem more temporary.

V.D. Management Spillovers

Given the evidence on the impact of these practices on firm performance, another key question is how much these practices spilled over to other firms. To address this in our ground-based survey (run from November 2011 to January 2012) we asked every one of the 96 nonproject firms if they had heard of the Stanford–World Bank project, and if so, what they knew (see Online Appendix AII for details). We found two results. First, the level of spillovers was extremely low, with only 16% of firms having heard of the project and only 2% having heard any details of actual practices. The reasons for the limited spillover were partly that owners were reluctant to discuss the details of their business with outsiders, and partly because they did not want to give their competitors information or risk them trying to hire away their plant managers. Second, to the extent that any spillovers occurred, these were entirely local—out of the five main textile towns around Mumbai, spillovers only arose in Tarapur and Umbergaon, where our experimental firms were located. As a result, although spillovers across plants within firms occurred rapidly as shown in Figure V, spillovers between firms were very limited.

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 nonadoption of any of the 38 practices in each plant. To do this consistently, we developed a flowchart (Online Appendix Exhibit OV), which runs through a series of questions to understand the root cause for the nonadoption 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 flowchart works, imagine a plant that does not record quality defects. The consultant would first ask if there was some external constraint, like labor regulations, preventing this, which we found never to be the case.[31] 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, although these plants' quality was often relatively good compared to other low-quality Indian textile plants, it was poor by international standards. So, in this case, the reason for nonadoption would be "incorrect information" because the owner appeared to have incorrect information on the cost-benefit calculation.

31. This does not mean labor regulations do not matter for some practices—for example, firing underperforming employees—but they did not directly impinge on the immediate adoption of the 38 practices.

The overall results for nonadoption of management practices are tabulated at two-month intervals starting the month before the intervention in Table IV. The rows report the different reasons for nonadoption as a percentage of all practices. These are split into nonadoption reasons for *common* practices (those that 50% or more of the plants were using before the experiment, like quality and inventory recording or worker bonuses) and *uncommon* practices (those that less than 5% of the plants were using in advance, like quality and inventory review meetings or manager bonuses). From Table IV, 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 nonadoption 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 the Japanese-inspired lean manufacturing revolution and are now standard across North America, Japan, and northern Europe but not in developing countries.[32]

Third, as the intervention progressed, the lack of information constraint on the uncommon practices was rapidly overcome in both treatment and control firms. It was easy to explain the existence of the uncommon management practices, so the nonadoption 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 percentage points).

Fourth, the incorrect information constraints were harder to address 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

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

TABLE IV

REASONS FOR THE NONADOPTION OF THE 38 MANAGEMENT PRACTICES (% OF ALL PRACTICES), BEFORE AND AFTER TREATMENT

Nonadoption 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	Treatment	Common	3.3	3.2	0.5	0	0	0
never heard of the practice	Treatment	Uncommon	64.0	19.1	2.9	1.5	0	0
before)	Control	Common	1.9	0	0	0	0	0
	Control	Uncommon	67.8	23.7	22.0	22.0	22.0	22.0
Incorrect information (heard of	Treatment	Common	30	22.4	15.4	15.2	14.4	14.4
the practice before but think	Treatment	Uncommon	30.9	50.7	50.7	49.3	49.3	47.1
it is not worth doing)	Control	Common	18.5	18.5	18.5	18.5	18.5	18.5
	Control	Uncommon	27.1	52.5	50.9	50.9	49.2	49.2
Owner time, ability, or pro-	Treatment	Common	1.1	0.8	0.5	0.8	1.6	0.8
crastination (the owner is	Treatment	Uncommon	3.7	13.2	13.2	13.2	13.2	14.0
the reason for nonadoption)	Control	Common	3.7	3.7	3.7	3.7	3.7	3.7
	Control	Uncommon	3.4	20.3	18.6	18.6	18.6	18.6
Other (variety of other	Treatment	Common	0	0	0	0	0	0
reasons)	Treatment	Uncommon	2.1	1.5	1.5	2.2	2.2	2.2
	Control	Common	0	0	0	0	0	0
	Control	Uncommon	0	0	0	0	0	0
Total nonadoption	Treatment	Common	34.6	26.4	16.3	16.0	16.0	15.2
	Treatment	Uncommon	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	Uncommon	98.3	96.6	91.5	91.5	89.8	89.8

Notes. Percentages (%) of practices not adopted by reason. Common practices are the eight practices with more than 50% initial adoption, mainly quality and downtime recording, and worker bonuses (see Appendix Table A.I for details). Uncommon practices are the 10 practices with less than 5% initial adoption, mainly standard operating procedures, inventory norms, 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). Nonadoption was monitored every other month using the tool shown in Online Appendix Exhibit OV, based on regular discussions with the firms' directors, managers, and workers.

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. As the consultants demonstrated the positive effect of these initial practice changes, the owners increasingly trusted them and adopted more of the recommendations, like performance incentives for managers. Thus, the common practice nonadoption 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 15.2% after nine months (a drop of 19.4 percentage points).

Fifth, 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 do so. A major reason is that the owners were severely time constrained, working an average of 68 hours per week already.[33] Although initially owners' time accounted for only 3.7% of nonadoption in treatment plants, by nine 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 effect of capital constraints, which are a significant obstacle to the expansion of micro-enterprises (e.g., De Mel, McKenzie, and Woodruff 2008). Our evidence suggested that these large firms were not cash-constrained, at least for tangible investments. We collected data on all the investments made by our 17 project firms during the two years of data collection. The mean (median) investment was \$880,000 (\$140,000). So investments on the scale of \$3,000 (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, although financial constraints do not appear to directly block the implementation of better management practices, they may hinder firms' ability

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

to improve their management using external consultants. Nevertheless, in conversations with factory managers, inability to borrow to finance consulting never came up as a reason for not using them, suggesting this is not the (main) binding constraint.

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 be to argue that poor management can at most be a short-run problem, because in the long run better-managed firms should take over the market. Yet most of our firms have been in business for more than 20 years.

One reason better run firms do not dominate the market appears to be constraints on growth derived from limited managerial span of control. In every firm in our sample, before the treatment, only members of the owning family had positions with any real decision-making power over finance, purchasing, operations, or employment. Non-family members were given only lower-level managerial positions with authority only over basic day-to-day activities. The principal reason seems to be that family members did not trust non-family members. For example, they were concerned if they let their plant managers procure yarn they may do so at inflated rates from friends and receive kickbacks.[34]

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 recover the assets is likely minimal because of the inefficiency of Indian courts. A compounding reason for the inability to decentralize in Indian firms seems to be the prevalence of bad management practices, as this meant the owners could not keep good track of materials and finance, and so might not even be able to identify mismanagement or theft within their firms.[35] This is consistent with the general finding of Bloom, Sadun, and van Reenen (2012)

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

35. 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

across a range of countries that firms headquartered in high-trust regions are more likely to decentralize, as are firms in countries with better rule of law.

As a result of this inability to delegate, firms could likely expand beyond the size that could be managed by a single director only if other family members were available to share executive responsibilities. Thus, as we saw in Table III, an important predictor of firm size was the number of male family members of the owners. 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 likely able to survive because more productive firms cannot expand. Our finding that improved management practice was connected with increased delegation to plant managers and investment in new plants supports this view.[36]

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—software, construction, and real estate—are growing rapidly, the attractiveness of new investment in textile manufacturing is relatively limited. Finally, a 35% tariff on cotton fabric imports insulates Indian textile firms against 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, so reallocation does not drive out badly run firms. Because entry is limited, new firms do not enter rapidly. The situation approximates a Melitz (2003)–style model with firms experiencing high decreasing

founding families place a huge emphasis on development and training (see also Banerjee and Duflo 2005).

36. Powell (2012) builds the intuition that the inability of owners in developing countries to trust nonfamily managers limits firm size into a contracting model and shows this explains a wide number of firm and macro facts.

returns to scale due to Lucas (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.[37]

The idea that firms might be run in a highly inefficient manner traces back at least to Leibenstein (1966). The management literature on the reasons for such inefficiency has tended to focus on three main drivers (e.g., Rivkin 2000; Gibbons and Henderson 2011). The first is the *motivation* problem (people know what would improve performance but lack the incentives) and is related to our limited competition and lack of delegation stories. The second and third match our informational story: the *inspiration* problem (the decision makers know they are not efficient but do not know how to fix this) and the *perception* problem (the firm's decision makers do not realize they are inefficient).

Of course, one question around informational stories is why firms do not learn, in particular from random variations in their operating conditions. For example, if the factory floor was less cluttered at some times than others, this would generate data that would suggest the value of neatness. Recent work by Schwartzstein (2012) and Hanna, Mullainathan, and Schwartzstein (2012) highlights, however, that cognitive limitations could prevent directors and managers from noting the impact of such variables that they had not thought important, so the information would be ignored. Certainly, our firms were sufficiently large and complex that the directors had to focus on certain areas, and we see this choice over what to focus on as an important area for future management research.

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 IV. Of course, consulting firms could approach firms for business, pointing out that their practices were

37. 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.

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 thus are unlikely to be receptive. Maybe 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' reluctance to reveal profits. Moreover, the client firm in such an arrangement might worry that the consultancy would twist its efforts to increase short-term profits in which it would share 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 17% within the first year from improved quality and efficiency and reduced inventory and appear to have been followed by a longer-run increase in firm size.

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, thereby, interfirm reallocation. Firms had not adopted these management practices before because of informational constraints. In particular, for many of the more widespread practices, although they had heard of these before, they were skeptical of their impact. For less common management practices, they simply had not heard of them.

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, multiplant 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, McKenzie, and Woodruff (2008)—may be simple enough that the owner can directly observe the full production process and hence does not need formal monitoring systems. Other interesting extensions involve examining in more detail 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. Our results also 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, something we also hope to explore in subsequent work.

Supplementary Material

An Online Appendix for this article can be found at QJE online (qje.oxfordjournals.org).

APPENDIX: SMALL SAMPLE TESTS SUMMARY

1. *Permutation Tests*

Permutation procedures use the fact that order statistics are sufficient and complete to propose and derive critical values for test statistics. We first implemented this for the null hypothesis of no treatment effect for the ITT parameter. This calculates the ITT coefficient for every possible combination of 11 treatment firms out of our 17 project 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 Online Appendix B).

APPENDIX TABLE A.I

THE TEXTILE MANAGEMENT PRACTICES ADOPTION RATES

Area	Specific practice	Preintervention level		Postintervention 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	0.667	0.071	0.167
	Scientific methods are used to define inventory norms for spares	0	0	0.071	0

APPENDIX TABLE A.I

(CONTINUED)

Area	Specific practice	Preintervention level		Postintervention change	
		Treatment	Control	Treatment	Control
Quality control	Quality defects are recorded*	0.929	1	0.071	0
	Quality defects are recorded defectwise	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

APPENDIX TABLE A.I

(CONTINUED)

Area	Specific practice	Preintervention level		Postintervention change	
		Treatment	Control	Treatment	Control
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 nonmanagerial 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 (August 2010) the management intervention. The practices with an asterisk were those also collected in the 2011 survey (see Online Appendix AI). The columns under Preintervention level of adoption report the preintervention share of plants adopting this practice for the 14 treatment and 6 control plants. The columns under Postintervention change in adoption report the changes in adoption rates between the preintervention period and four months after the end of the diagnostic phase (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.

APPENDIX TABLE A.II

OLS and IV Estimations of the Effect of Management Practices on Plant Performance

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
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
Management$_{i,t}$	-0.558	-1.694**	-0.404	-0.833**	0.119	0.310**	0.167	0.477**
	(0.440)	(0.850)	(0.259)	(0.347)	(0.099)	(0.128)	(0.176)	(0.242)
Specification		IV 1st stage		IV 1st stage		IV 1st stage		IV 1st stage
Dependent variable		Management		Management		Management		Management
Cumulative treatment$_{i,t}$		0.018***		0.017***		0.019***		0.019***
		(0.002)		(0.002)		(0.002)		(0.002)
Small sample robustness								
Ibragimov-Mueller (95% CI)	[−19.00,5.22]	[−5.28,-1.18]	[−4.77,0.11]	[−1.50,-0.57]	[0.27,1.38]	[0.15,1.21]	[−3.59,11.97]	[0.27,1.90]
IV Permutation Tests (95% CI)		[−28.05,-0.18]		[−6.85,0.37]		[0.09,0.78]		[0.44,3.91]
First stage F-test		73.41		72.88		107.55		86.00
Time FEs	127	127	127	127	127	127	127	127
Plant FEs	20	20	18	18	20	20	20	20
Observations	1,807	1,807	2,052	2,052	2,393	2,393	1,831	1,831

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 Online Appendix Table AI. *Cumulative treatment* is the cumulative weeks of since beginning the implementation phase in each plant (0 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 are available. *Small sample robustness* implements three different procedures (described in greater detail in Online Appendix B) to address issues of plant heterogeneity, within plant (and firm) correlation, and small sample concerns, where *95% CI* reports 95% 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 1,000 possible permutations (out of 12,376) of treatment assignment. *IV-Permutation* tests implements a permutation test for the IV parameter using 1,000 possible permutations (out of 12,376) of treatment assignment. These tests have exact finite sample size. *** denotes 1%, ** denotes 5%, * denotes 10% significance.

2. 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 Ibramigov and Mueller (2010) to implement a t-statistic based estimator that is robust to substantial heterogeneity across firms as well as to autocorrelation across observations within a firm. This approach requires estimating the parameter of interest separately for each treatment and control firm and then comparing the average of the 11 treatment firm estimates to those of the 6 control firms using a standard t-test for grouped means (allowing for unequal variances) with 5 degrees of freedom (see Online Appendix B). Such a procedure is valid in the sense of having the correct size (for a fixed small number of firms) as long as the time dimension is large enough that the estimate for each firm can be treated as a draw from a normal distribution.

Stanford University, CEP, CEPR, NBER, and SIEPR
Overland Advisors
Stanford University and SCID
World Bank, BREAD, CEPR, and IZA
Stanford University and SIEPR

References

Andrews, Donald W. K., and Vadim Marmer, "Exactly Distribution-free Inference in Instrumental Variables Regression with Possibly Weak Instruments," *Journal of Econometrics*, 142, no. 1 (2008), 183–200.

Banerjee, Abhijit, and Esther Duflo, "Growth through the Lens of Development Economics," In *Handbook of Economic Growth, Volume 1*, ed. Philippe Aghion and Stephen Durlauf (Amsterdam: Elsevier North-Holland, 2005).

Bertrand, Marianne, and Antoinette Schoar, "Managing with Style: The Effects of Managers on Corporate Policy," *Quarterly Journal of Economics*, 118, no. 4 (2003), 1169–1208.

Black, Sandra, and Lisa Lynch, "What's Driving the New Economy? The Benefits of Workplace Innovation," *Economic Journal*, 114, no. 493 (2004), 97–116.

Bloom, N., G. Genakos, R. Martin, and R. Sadun, "Modern Management: Good for the Environment or Just Hot Air?" *Economic Journal*, 120, no. 544 (2010), 551–572.

Bloom, Nicholas, Raffaella Sadun, and John Van Reenen, "The Organization of Firms across Countries," *Quarterly Journal of Economics*, 127, no. 4 (2012), 1663–1705.

Bloom, Nicholas, and John Van Reenen, "Measuring and Explaining Management Practices across Firms and Countries," *Quarterly Journal of Economics*, 122, no. 4 (2007), 1341–1408.

———, "Why Do Management Practices Differ Across Firms and Countries?" *Journal of Economic Perspectives* (2010), 203–224.

——— "Human Resource Management and Productivity," In *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card (Amsterdam: Elsevier North-Holland, 2011).
Bruhn, Miriam, Dean Karlan, and Antoinette Schoar, "The Impact of Offering Consulting Services to Small and Medium Enterprises" (Mimeo, World Bank, 2012).
Bruhn, Miriam, and David McKenzie, "In Pursuit of Balance: Randomization in Practice in Development Field Experiments," *American Economic Journal: Applied Economics*, 1, no. 4 (2009), 200–232.
Bruhn, Miriam, and Bilal Zia, "Stimulating Managerial Capital in Emerging Markets: The Impact of Business and Financial Literacy for Young Entrepreneurs," World Bank Working Paper 5642, 2011.
Cameron, A. Colin, Jonah Gelbach, and Douglas Miller, "Bootstrap-Based Improvements for Inference with Clustered Errors," *Review of Economics and Statistics*, 90, no. 3 (2008), 414–427.
Cappelli, Peter, and David Neumark, "Do 'High Performance' Work Practices Improve Establishment-Level Outcomes?" *Industrial and Labor Relations Review*, 54, no. 4 (2001), 737–775.
Caselli, Francesco, "CREI Lectures in Growth" (Mimeo, LSE, 2011).
Caselli, Francesco, and Nicolas Gennaioli, "Dynastic Management," (Mimeo, LSE, 2011).
Clark, Greg, "Why Isn't the Whole World Developed? Lessons from the Cotton Mills," *Journal of Economic History*, 47 (1987), 141–173.
Conley, Tim, and Christopher Udry, "Learning about a New Technology: Pineapple in Ghana," *American Economic Review*, 100, no. 1 (2010), 35–69.
De Mel, Suresh, David McKenzie, and Christopher Woodruff, "Returns to Capital in Microenterprises: Evidence from a Field Experiment," *Quarterly Journal of Economics*, 113, no. 4 (2008), 1329–1372.
Drexler, Alejandro, Greg Fischer, and Antoinette Schoar, "Financial Literacy Training and Rule of Thumbs: Evidence from a Field Experiment," (Mimeo, MIT, 2010).
Duflo, Esther, Michael Kremer, and Jonathan Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," *American Economic Review*, 101, no. 6 (2011), 2350–2390.
Foster, Lucia, Haltiwanger John, and Chad Syverson, "Reallocation Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" *American Economic Review*, 98, no. 1 (2008), 394–425.
Gibbons, Robert, and Rebecca Henderson, "Relational Contracts and Organizational Capabilities," *Organization Science*, 23, no. 5 (2011), 1350–1364.
Greevy, Robert, Jeffrey H. Silber, Avital Cnaan, and Paul R. Rosenbaum, "Randomization Inference with Imperfect Compliance in the ACE-Inhibitor" *Journal of the American Statistical Association*, 99, no. 465 (2004), 7–15.
Hanna, Rema, Sendhil Mullainathan, and Josh Schwartzstein, "Learning through Noticing: Theory and Experimental Evidence in Farming" (Harvard: Mimeo, 2012).
Harrison, Glenn, Morten Lau, and Elisabeth Rustrom, "Risk Attitudes, Randomization to Treatment, and Self-Selection into Experiments," *Journal of Economic Behavior and Organization*, 70, no. 3 (2009), 498–507.
Heckman, James, "Randomization in Social Programs," In *Evaluating Welfare and Training Programs*, ed. C. Manski and I. Garfinkle (Cambridge MA: Harvard University Press, 1992).
Hsieh, Chiang-Tai, and Pete Klenow, "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124, no. 4 (2009), 1403–1448.
———, "Development Accounting," *American Economic Journal: Macroeconomics*, 2, no. 1 (2010), 207–223.
Ibragimov, Rustam, and Ulrich Muller, "*t*-statistic Based Correlation and Heterogeneity Robust Inference," *Journal of Business and Economic Statistics*, 28, no. 4 (2010), 453–468.
Ichniowski, Casey, Kathryn Shaw, and Giovanna Prennushi, "The Effects of Human Resource Management: A Study of Steel Finishing Lines," *American Economic Review*, 87, no. 3 (1997), 291–313.

Karlan, Dean, Ryan Knight, and Chris Udry, “Hoping to Win, Expected to Lose: Theory and Lessons on Microenterprise Development,” NBER working paper 18325, 2012.
Karlan, Dean, and Martin Valdivia, “Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions,” *Review of Economics and Statistics*, 93, no. 2 (2011), 510–527.
Lazear, Edward, Ulrike Malmendier, and Roberto Weber, “Sorting in Experiments with Application to Social Preferences,” *American Economic Journal: Applied Economics*, 4, no. 1 (2012), 136–163.
Lazear, Edward, and Paul Oyer “Personnel Economics,” In *Handbook of Organizational Economics*, ed. Robert Gibbons and John Roberts (Princeton: Princeton University Press, 2012).
Leibenstein, H., “Allocative Efficiency vs. X-efficiency,” *American Economic Review*, 56, no. 3 (1966), 392–415.
Locke, Richard, Fei Qin, and Alberto Brause, “Does Monitoring Improve Labor Standards? Lessons from Nike,” *Industrial and Labor Relations Review*, 61, no. 1 (2007), 3–31.
Lucas, Robert, “On the Size Distribution of Business Firms,” *Bell Journal of Economics* (1978), 508–523.
McKenzie, David, “Impact Assessments in Finance and Private Sector Development: What Have We Learned and What Should We Learn?” *World Bank Research Observer*, 25, no. 2 (2010), 209–233.
———, “Beyond Baseline and Follow-up: The Case for More *T* in Experiments,” *Journal of Development Economics* (2012), forthcoming.
McKenzie, David, and Christopher Woodruff, “What Are We Learning from Business Training and Entrepreneurship Evaluations around the Developing World?” (Mimeo, World Bank, 2012).
Melitz, Marc, “The Impact of Trade on Intra-industry Reallocations and Intra-industry Productivity,” *Econometrica*, 71, no. 6 (2003), 1695–1725.
Milgrom, Paul, and John Roberts, “The Economics of Modern Manufacturing: Technology, Strategy and Organization,” *American Economic Review*, 80, no. 3 (1990), 511–528.
Mundlak, Yair, “Empirical Production Function Free of Management Bias,” *Journal of Farm Economics*, 43, no. 1 (1961), 44–56.
Powell, Michael, “Productivity and Credulity in Industry Equilibrium” (Mimeo, MIT, 2012).
Rivkin, Jan, “Imitation of Complex Strategies,” *Management Science*, 46, no. 6 (2000), 824–844.
Schwartzstein, Josh, “Selective Attention and Learning,” Working Paper, Dartmouth, 2012.
Stigler, George, “The Xistence of X-efficiency,” *American Economic Review*, 66, no. 1 (1976), 213–216.
Syverson, Chad, “Market Structure and Productivity: A Concrete Example,” *Journal of Political Economy*, 112, no. 6 (2004), 1181–1222.
———, “What Determines Productivity at the Micro Level?” *Journal of Economic Literature*, 49, no. 2 (2011), 326–365.
Suri, Tavneet, “Selection and Comparative Advantage in Technology Adoption,” *Econometrica*, 79, no. 1 (2011), 159–209.
Taylor, Frederick. *Principles of Scientific Management* (New York: Harper and Brothers, 1911).
Walker, Francis, “The Source of Business Profits,” *Quarterly Journal of Economics*, 1, no. 3 (1887), 265–288.
Womack, James, Daniel Jones, and Daniel Roos, *The Machine that Changed the World* (New York: Simon and Schuster, 1990).
Woodward, John, *Management and Technology* (Cambridge: Cambridge University Press, 1958)