ON-LINE APPENDIX A: DATA

AI) Performance Data

Our estimates for profits are laid out in On-line Appendix Table OI, with the methodology outlined below. We calculate the numbers for the median firm. We first generate the estimated impacts on quality, inventory, and efficiency using the Intention to Treat (ITT) numbers from Table II, which shows a reduction of quality defects of 43.2% (exp(-0.564)-1), a reduction in inventory of 21.7% (exp(-0.245)-1) and an increase in output of 9.4% (exp(0.090)-1).

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

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

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

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

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

AII) Management survey in 2011

In 2011 we decided to run a broad industry level survey to collect data on all medium (100 to 1000 employee) textile weaving firms around Mumbai. We wanted to collect information on the wider textile industry in the Mumbai region to compare against our project firms.

We started the survey process by building a population database of every textile firm with 100 to 1000 employees around Mumbai[1]. This came from the Ministry of Commercial Affairs (MCA) registry of firms, plus industry association lists, internet searches, yellow pages and telephone directories and lists of

[1] "Around Mumbai" we define as firms for which Mumbai is the natural headquarter location. This covered all of Maharashtra, southern Gujarat (towns that are closer to Mumbai than Ahmedabad) and Dadra and Nagar Haveli.

firms provided by our field experiment firms and their suppliers and customers. This helped to supplement the official MCA list as many textile firms had been incorrectly allocated to other industries (like spinning, processing or apparel) so had not shown up on our MCA original textile list. Through this process, we identified 172 firms meeting the size (100 to 1000 employees), industry (cotton weaving) and location (around Mumbai) criteria.

We then started by telephoning every firm on this list from Stanford in October 2011. These calls were initiated by Aprajit Mahajan, who introduced himself as "Professor Mahajan calling from Stanford, USA" to emphasize the research nature of the project. All firms that showed any potential interest in the survey were then sent by Federal Express from Stanford a box containing Stanford and World Bank publications and Stanford clothing (two t-shirts and a cap). One week later the firms were then telephoned (often on multiple occasions) by our international consulting firm from Mumbai to arrange a face-to-face interview with a Director (typically the CEO). These steps were important because the Directors of these firms are extremely busy, and many are also suspicious of outside organizations. Initiating the call from the USA (using a number with international caller-ID) and sending Federal Express packages from Stanford helped to emphasize this was a legitimate international research exercise. Having consultants from a high-profile international firm arrange and run the interviews face-to-face highlighted the importance of the research.

These interviews were conducted over a 10 week period between November 2011 and January 2012. We obtained data from 113 firms (66% of the full sample), including our 17 project firms and 96 non-project firms. We interviewed our experimental firms in addition to the new firms to ensure we had comparable data on them. In addition, we recorded 8 firms that had exited the industry during this time, and we include these in our plant size regressions, conservatively entering them as having one plant in 2008 and zero plants at the end of 2011.

The surveys themselves took place usually at one of the firm's plant (88%) or at their Mumbai headquarters (12%). They lasted, on average, 49 minutes and were carried in English (54%) and Hindi (46%). They were carried out by a team of three consultants, and for internal consistency 30% of the interviews had two consultants attending and 6% had all three consultants attending. A series of interview noise controls were also collected after the interview following Bloom and Van Reenen (2007), like time, date, duration, respondent characteristics (age, education), and a self-assessed reliability score. We also double interviewed 15 firms by returning to run a second interview on another Director.

The interview[2] followed a relatively standardized script, asking background questions about the firm (age, ownership, family involvement, markets etc), followed by questions about plant size (employees, output, plant numbers, production quantity), management practices, organizational structure, computerization, prior consulting, prior knowledge of the Stanford-World Bank project (we skipped this question for firms involved in the experiment), and any potential interest in future consulting waves. The full survey is available at www.stanford.edu/~nbloom/Template.xlsx. Because of the time limit, we could only ask questions on 16 of the 38 management practices. These 16 practices have been starred in Table AI, and were selected as the questions that were easiest to collect accurate data on in an interview and informative about overall management practices. They covered mainly preventive maintenance, quality control and inventory management. The average score across these 16 questions had a correlation of 0.951 with the 38 questions in our 28 plants, suggesting these 16 questions are an extremely good guide to firms overall management scores. So, to generate management scores for our survey firms for Figure II (the management adoption figure), we extrapolated the scores for all 38 questions based on the 16 questions in the survey.

[2] The full interview is available here http://www.stanford.edu/~nbloom/Template.xls

For variables that could change over time (like management and the number of plants) we asked information for four dates: August 1st 2008, December 1st 2009, December 1st 2010 and December 1st 2011. The first date was chosen to be just before the beginning of the experiment while the other dates were chosen to coincide with the year end (and for 2011 with the interview timing).

This interview process was extremely expensive ($150,000 for the consultants and $10,000 for the FedEx pages). The high cost of the consultants was because the team of three consultants ran on average 12.8 interviews a week (less than 1 per person per day). This was because of the extensive travel times between factories and also the frequent need to reschedule interviews (the consultants would often travel to meet the CEO at his factory to find him not present). However, it generated an extremely high response rate considering the target population of wealthy Directors of large (100 to 1000 employee) firms. Our estimate is that most all of these were millionaires and a few of them were worth several hundred million dollars (from other business interests and land holding in Mumbai).

AIII) Measuring Productivity

We define productivity=log(value added) – 0.42*log(capital) – 0.58*log(labor). The factor weights are the cost shares for cotton-weaving in the Indian Annual Survey of Industry (2004-05). The output and input measures are obtained as follows:

Value-Added: Log(value-added$_{i,t}$)=log(output$_{i,t}$) + f_i where f_i is a plant fixed effect (which drops out in all estimations and plots since we always examine changes over time). This approach works because output is measured in terms of production picks – a physical concept that is the number of cycles of the weaving shuttle. Because each production pick requires a constant amount of weft yarn, warp yarn and electricity to operate the loom-shuttle, material inputs are proportional to output. So, changes in log(value-added) are equal changes in log(output) within each plant.

Capital: This includes all land, buildings, equipment, and inventory in the plant. The first three components (land, buildings and equipment) were constant over the experimental period (January 2008 until August 2010) as we focused on a fixed set of looms in each plant, and the plants did not change these looms over this period. The fourth component – inventory – does change, and we measured this over time. We combined all four terms by current market value to create a composite capital measure (our consulting firm provided estimates of the value of the land, building and equipment for each firm from local factory and equipment resale prices).

Labor: This included all labor employed by the firms (the managers) and labor employed by contractors (weavers and mending labor). This was evaluated in terms of hours where managers were assumed to work the standard shift (6 days a week for 12 hours) and weavers and mending labor the shifts they were contracted for (typically 28 days a month for 12 hours a day, except for plants which employed female menders who worked 8 hour shifts to enable daylight commutes).

AIV) How we gathered firm level data

Data collection for the performance metrics was undertaken on a plant by plant basis. In every plant, the consultants worked with the firms to collect a quality defects index (QDI), output (production picks), inventory (tons of yarn), and workers (numbers and/or hours). The QDI and output data was usually collected daily, while for workers and inventory the frequency was weekly as payroll and inventory tallies are calculated weekly. Prior to the diagnostic phase, firms typically collected data but did not examine trends in it or have systems in place to act upon it.

Moreover, records were often in paper form or in enterprise resource planning (ERP) systems whose full functionality was not being taken advantage of. The diagnostic phase therefore involved constructing historic series for all the data metrics on the basis of written log books, extracts from these ERPs, order forms, and other firm records. They also put in place systems for easier use of this data going forward. The intervention then affected how this data was used, rather than the underlying measurement of data.

For the management variables, data was collected through a combination of direct observation and interviews with the plant managers and Directors. For example, practices like "quality defects are recorded" and "quality defects are recorded defectwise" are easy to observe, while practices like "There is a reward system for managerial staff based on performance" requires asking plant managers and Directors. Historical values for management variables were collected by asking managers and Directors about these practices on the first of each month for the historic month. This was reasonably easy to do because few of these practices change, and because they are very specific it is easy for managers to recollect if and when they changed.

ON-LINE APPENDIX B: ECONOMETRICS

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

The proposed procedure by Ibragimov-Mueller (2009, 2012) (IM) is useful for our case where the number of entities (firms) is small but the number of observations per entity is large. Their approach can be summarized as follows: Implement the estimation method (OLS, IV, ITT) on each firm separately and obtain a set of 17 firm-specific estimates.[3] Then compare 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. Note that we cannot do this for the IV estimand since we cannot implement an IV procedure on any of the control firms alone. In this case, we compute the IV estimate for each treatment firm and treat the resulting estimates as draws from a t-distribution with 10 (N-1 of the 11 treatment firms) degrees of freedom (see IM 2009). The results from this procedure are based on before-after comparisons for the treatment firms (using the control firms to remove time period effects).

The procedure requires that the coefficient estimates from each entity are asymptotically independent and Gaussian (but can have different variances). In our case, this would be justified by an asymptotics in T argument (recall we have over a 100 observations per plant). In particular, we can be agnostic about the exact structure of correlations between observations within a firm as long as the parameter estimators satisfy a central limit theorem. Subject to this requirement, the extent of correlation across observations within an entity is unrestricted. In

[3] To be consistent with our main results in Table II we estimate the specification (1) for each firm. We note though that given the form of the two-sample test we do not need to estimate the time-effects for the IM procedure (at least over period where all firms are observed). Results from this procedure are substantively similar to those reported here and are available upon request.

addition, different correlation structures across firms are permissible since the procedure allows for different variances for each firm level parameter. This "asymptotic heterogeneity" considerably relaxes the usual assumptions made in standard panel data contexts (such as those underlying the cluster covariance matrices in our main tables). Finally, IM show that the limiting standard Gaussian distribution assumption (for each firm) can be relaxed to accommodate heterogeneous scale mixtures of standard normal distributions as well. The asymptotic arguments imply that we can treat the firm-by-firm estimates as draws from independent normal distributions and we use this to conduct inference. Note that this procedure works (i.e. the tests have correct asymptotic size) even though the observations are heterogeneous in that they have variances (and potentially different within-firm correlation structures).

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

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

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

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

Under the null hypothesis of no treatment effect, the treatment outcomes should not be systematically larger than the control outcomes. Specifically, under the null hypothesis and conditional upon the order statistics, each possible candidate value of T has an equal probability of occurring. We use this insight to construct a critical value for the test. Consider one of the $\binom{17}{11}$ combinations of the firm treatment assignment variable *Z*. For each such permutation, compute *T*. Form the empirical distribution of T by considering all possible permutations and record the appropriate quantile for the distribution of *T* thus generated (in the one-sided alternative case this would be the $1-\alpha$ quantile). Finally, reject the null hypothesis of no treatment effect if the original statistic *T* exceeds this quantile. Greevy et al (2004), show that this test has exact size α for any sample size *n*. Therefore, the conclusions of this test do not rely upon any asymptotic theory. Instead, the results lean heavily on the assumption of exchangeability – the property that changing the ordering of a sequence of random variables does not affect their joint distribution. For our application, this notion seems reasonable. Note that exchangeability is weaker than the i.i.d. assumption so for instance outcomes across firms can even be correlated (as long as they are equi-correlated) and also we do not require that observations within a firm are independent over time.

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

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

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

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

and we form the statistic as

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

Where

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

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

ON-LINE APPENDIX C: FIXED-EFFECT VS. IV ESTIMATIONS OF THE PERFORMANCE IMPACT OF MANAGEMENT

A growing number of papers estimate the impact of management practices on firm and plant performance by running OLS fixed effects regressions of the type:[4]

$$\mathrm{OUTCOME}_{i,t} = \alpha_i + \beta_t + \theta \mathrm{MANAGEMENT}_{i,t} + \nu_{i,t} \tag{2}$$

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

The concern is that changes in management practices are not exogenous to changes in the outcomes that are being assessed, so that the coefficient θ on management could be biased.

Our study provides an opportunity to examine this by comparing fixed effects coefficients with the experimentally identified IV coefficients. To do this, we instrument the management practice score with cumulative weeks of the intervention treatment. The exclusion restriction is that the intervention affected the outcome of interest only through its impact on management practices, and not through any other channel. A justification for this assumption is that the consulting firm focused entirely on the 38 management practices in their recommendations to firms, and firms did not hire new labor and made only trivial investments as a result of the intervention during the period of our data (at least until August 2010). Nevertheless, we acknowledge that it is possible that the management consultants may have made suggestions that impacted on outcomes through channels other than the 38 basic management practices, which would cause this exclusion restriction to be violated. However, we did not hear of any such cases when directly asking firm owners what the main things they had learned from the consultants were.

We see in Table AII that the fixed effects estimate for the impact of management practices are less than one-half the IV estimates. For example, the fixed-effects impact of adoption of management practices on TFP is 0.242 compared to the IV coefficient of 0.523. One possible reason for this heavy downward bias is measurement error in the management variable, causing attenuation bias in our fixed effects estimates. However, our management practice measures are binary indicators that are collected every other month by the consultants, so they should be accurately measured. From discussions with the consultants and owners, it appears instead that the main reason for this downward bias is that plants were more willing to adopt new management practices when performance was deteriorating compared to when it was stable or improving. This is consistent with a long stream of evidence suggesting that bad times spur reorganizations (see, for example, Leibenstein 1966).

Table OI: The impact of modern management practices on plant performance: robustness to controls for *relative* timing

Dependent Variable	Quality defects	Inventory	Output	TFP	Quality defects	Inventory	Output	TFP
Specification	ITT	ITT	ITT	ITT	Weeks of Treatment	Weeks of Treatment	Weeks of Treatment	Weeks of Treatment
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$\text{Intervention}_{i,t}$	-0.659*	-0.292**	0.119**	0.193**				
Post implementation stage	(0.382)	(0.134)	(0.049)	(0.089)				
During Implementation	-0.389***	-0.061	0.032	0.139***				
During the implementation stage	(0.136)	(0.116)	(0.025)	(0.041)				
$\text{Cumulative treatment}_{i,t}$					-0.038*	-0.016**	0.007***	0.011**
Total weeks of implementation					(0.021)	(0.007)	(0.003)	(0.005)
Time FEs	127	127	127	127	127	127	127	127
Plant FEs	20	18	20	18	20	18	20	18
Observations	1807	2052	2393	1831	1807	2052	2393	1831

Notes: All regressions use a full set of plant and weeks *relative* to the start of the diagnostic phase dummies (rather than calendar weeks as in Table II). Standard errors are bootstrap clustered at the firm level. **Intervention** is a plant level dummy equal to one after the implementation phase at treatment plants and zero otherwise. **During Implementation** is a dummy variable equal to one six months from the beginning of the diagnostic phase for all treatment plants. **Cumulative treatment** is the cumulative weeks of treatment since the beginning of the implementation phase in each plant (zero 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 is available. *** denotes 1%, ** denotes 5%, * denotes 10%.

Table OII: Estimated median impact on profits

Change	Impact	Estimation approach	Estimated impact
Improvement in quality	Reduction in repair manpower	Reduction in defects (43%) times median mending manpower wage bill ($41,000).	$18,000
	Reduction in waste fabric	Reduction in defects (43%) times the average yearly waste fabric (5%) times median average sales ($6m).	$129,000
Reduction in inventory	Reduction in inventory carrying costs	Reduction in inventory (22%) times carrying cost of inventory (22%) times median inventory ($230,000)	$11,000
Increased efficiency	Increased sales	Increase in output (9%) times margin on sales (31%) times median sales ($6m)	$167,000
Total			$325,000

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