The Dynamics and Spillovers of Management Interventions: Evidence from the Training Within Industry Program

Nicola Bianchi and Michela Giorcelli*

July 9, 2020

Abstract: This paper examines the effects of management training on firm performance, using evidence from the Training Within Industry (TWI) program. The TWI plan was implemented by the U.S. government between 1940 and 1945 to provide management training to firms involved in war production. Using newly collected panel data on all 11,575 U.S. firms that applied to the program, we find that the TWI training had a positive and long-lasting effect on firm performance. Trained firms sustained better performances in the long run through the adoption of better practices, complementarities among training fields, and positive spillover effects in their supply chain. (*JEL:* L2, M2, N34, N64, O32, O33)

*Bianchi: Kellogg School of Management, Northwestern University, 2211 Campus Drive, Evanston, IL 60208, and NBER, nicola.bianchi@kellogg.northwestern.edu; Giorcelli: Department of Economics, University of California - Los Angeles, 9262 Bunche Hall, Los Angeles, CA 90095, and NBER, email: mgiorcelli@econ.ucla.edu. We thank Philipp Ager, Andy Atkenson, Simon Board, Ryan Boone, Bruno Caprettini, Dora Costa, Alessandra Fenizia, Mitch Hoffman, Giampaolo Lecce, Claudia Martinez, Niko Matouschek, Joel Mokyr, Adriana Lleras-Muney, Giuseppe Nicoletti, Jagadeesh Sivadasan, Melanie Wasserman, as well as seminar and conference participants at Bocconi, Bologna, Chicago Booth, HBS, MIT, Monash University, Mountain West Economic History Conference, NYU Stern, OECD Productivity Forum, PUC-Rio de Janeiro, PUC-Santiago, Queen's Smith School of Business, UBC, UCLA, UCSD, Universidad Torcuato Di Tella, University of New South Wales, University of Queensland, University of Sydney, and Yale. We are grateful to Dmitri Koustas and James Lee for kindly sharing with us the war supply contracts data and the 1939 Census data, respectively. Juan Rojas and Zhihao Xu provided outstanding research assistance.

1 Introduction

Economists have long speculated on the reasons behind the large and persistent variations in productivity across establishments in both developed and developing countries (Syverson, 2004; Foster, Haltiwanger, and Syverson, 2008; Hsieh and Klenow, 2009). A possible explanation could be ascribed to differences in the adoption of managerial practices (Ichniowski, Shaw, and Prennushi, 1997; Bloom and Van Reenen, 2007). Consistent with this view, more recent papers have shown that management has causal effects on firm performance, confirming its role in driving differences in productivity (Bloom et al., 2013; Bruhn, Dean, and Schoar, 2018; Cai and Szeidl, 2017; Giorcelli, 2019).

However, most of the available evidence comes from relatively small-scale randomized controlled trials (RCTs) that measure the effects of better management only in the short run, usually covering a few months after the original intervention. Therefore, little is known about the long-term dynamics of management interventions on firm outcomes. Providing more evidence on this topic is important because extrapolating the long-term effect of better management from short-term estimates is not an easy task.

On the one hand, the so-called "Toyota-way" hypothesis states that a single management intervention can put firms on a virtuous cycle of continuous improvements, also know as the *kaizen* approach (Liker, 2004). For example, after an initial positive shock to management, complementarities among managerial areas or positive spillovers along the supply chain could make it easier for firms to get better over time (Womack, Jones, and Roos, 1990). On the other hand, the "anti-*kaizen*" view states that the beneficial effects of management interventions are not sustainable over time. After an initial boost to firm performance, turnover of good managers and changes in market conditions could lead firms to abandon the adoption of good practices in the long run (Karlan, Knight, and Udry, 2015; Higuchi, Mhede, and Sonobe, 2016; Bloom et al., 2020).

This paper studies the long-term dynamics of better management, using evidence from a unique historical episode, the Training Within Industry (TWI) program. The TWI program was implemented by the U.S. government between 1940 and 1945 with the purpose of providing management training to firms involved in war production. It reached 11,575 large U.S. firms across different economic sectors and geographical areas. The program offered in-plant training in three main areas, called J modules: Job-Instructions (J-I), Job-Relations (J-R), and Job-Methods (J-M). The J-I module taught supervisors and managers how to train new workers, the J-R module how to promote teamwork, and the J-M module how to introduce improvements to current processes.

We use newly collected panel data on the population of 11,575 U.S. firms that applied to the TWI program. For each firm, we collected balance sheets and income statements from 1935 to 1955. We matched this financial information to data on the TWI program that we digitized from the TWI program's historical records.

Our empirical strategy measures the effects of management training by comparing over time applicant firms that received the training to applicant firms that did not receive it, controlling for application date, county and sector. A plausible concern is that firms with and without training were fundamentally different in ways that may bias the analysis. For example, the TWI administration might have preferred to focus its limited staff on companies that were already more productive or on a faster trajectory of growth at baseline, on those that were located in geographical areas that were more important for the military, or those that were operating in certain sectors. Had this form of selection into training happened, comparing firms with and without training could confound the effects of the TWI program with other unrelated factors.

However, the data indicate that firms with and without training had statistically similar characteristics at baseline and were located in statistically similar counties. Moreover, the probability of being trained in a given location was not autocorrelated, indicating that the TWI administration did not favor over time firms located in certain regions. Finally, and more importantly for our difference-in-differences specification, we show that firms with and without training were following statistically indistinguishable trends in performance before the start of the TWI program.

In short, we compare similar firms that applied to the program on the same day from nearby locations within the same county. The fact that only some were trained was driven by idiosyncrasies in the implementation of the TWI program. Applicant firms were divided into smaller geographical groups, called subdistricts. The in-plant training was then carried out by TWI instructors, whom the TWI administration trained to teach one J-module before assigning them to a given subdistrict. The lack of sufficient manpower and the inability of TWI instructors to teach in multiple subdistricts or more than one module created imbalances in the composition of instructors across subdistricts and over time. As a result, some applicants received training only in some J-modules, while others did not receive any training at all.

Our analysis starts by estimating the long-term dynamics of the TWI training over ten years after the program. Overall, we find that the effects were positive and long-lasting. For example, sales of trained firms increased by 5.3 percent within one year of the TWI training, compared with nontrained applicants. This effect peaked at 21.7 percent in period eight and then decreased to 16 percent in period ten. The effects on productivity and profitability increased

monotonically throughout the post-training period.

After assessing the overall effect of the TWI program, we start opening the black box of training by establishing a link between the content of the J-module and the adoption of different managerial practices. For this analysis, we take advantage of the fact that different firms received different J-modules, due to the aforementioned incomplete assignment of instructors to different subdistricts. The data indicate that, although all trained applicants experienced increases in performance, firms with different training implemented different changes to their organization to achieve higher productivity. For example, firms trained in J-R learned how to motivate workers and resolve internal disputes. Consistent with their training, their financial reports indicate that they increased spending for performance-based bonuses, became more likely to report spending for on-the-job training, and experienced fewer strikes. Moreover, these firms did not report changes in practices not related to the specific training they had received, reinforcing the link between the content of the J-modules and firm outcomes.

Overall, our findings support the "Toyota-way" hypothesis. Trained firms were able to improve their performance for many years after training. Moreover, they experienced long-lasting changes in managerial practices. Therefore, the second part of the analysis aims at exploring the channels that might have allowed trained firms to achieve better outcomes in the long run. We establish two additional results. First, there were complementarities in the J-R and J-M modules. For example, firms that received the J-R training with another module experienced fewer strikes and were more likely to report spending for on-the-job training than firms that received the J-R module by itself. This finding is important because it is one of the first empirical tests of a core component of the "Toyota-way" hypothesis. Due to complementarities, the effects of one type of training can generate spillovers on other fields, making it easier for firms to get better over time.

Second, the TWI program had positive effects on the supply chain of trained firms. Specifically, after applicants received the training, their upstream and downstream firms started becoming more productive. Remarkably, each J-module induced upstream and downstream firms to change different practices. For example, firms in the supply chain of applicants that received the J-R training became less likely to experience worker strikes and more likely to report spending for on-the-job training, mimicking the same changes implemented by the trained firm. Moreover, trained applicants became more likely to select better upstream and downstream firms after receiving training. These findings suggest another channel through which the program might have allowed trained firms to achieve *kaizen* or long-lasting improvement. In fact, the TWI training did not limit its influence within the boundaries of the trained firms, but it slowly extended over time along the supply chain finding new opportunities for growth.

On the contrary, we do not find evidence of negative horizontal spillovers. Firms eligible to participate in the TWI program that did not apply but were geographically close to trained firms did not experience negative effects on their financial outcomes. These results do not necessarily rule out the existence of business stealing, because the U.S. economy was growing during the post-WWII era. In other words, trained firms might have been able to gain market shares over nonapplicants even without poaching their current clients.

Other explanations that have been advanced against the "Toyota-way" hypothesis did not seem to apply to our sample. For example, firms with training were able to increase their productivity even in the face of high turnover of top managers after WWII. This result informs the debate on whether the benefits of management interventions belong primarily to firms or the trained managers. In our case, the TWI program had a strong firm-specific component that stayed within the firm even after many trained managers left the company.

Finally, other war-related factors did not play a big role in determining the effect of training. For example, the observed increase in productivity and sales is not driven by more war contracts. In other words, firms with training did not improve simply because they gained a tighter relationship with the federal government. Moreover, exposure to the draft and the disruptions associated with switching between peacetime and war production were not able to erase the benefits of the TWI program.

The contribution of this paper is twofold. First, recent studies in the literature on management have provided causal evidence on the relationship between management interventions and firm performance. The results of these experiments tend to vary depending on the nature of the intervention. While RCTs that provide management consulting usually show large and positive results (for example, Bloom et al., 2013; Bruhn, Dean, and Schoar, 2018), the effects of management training tend to be more mixed (McKenzie and Woodruff, 2014). The TWI program was primarily a management training scheme, but had a consulting component. Although the lectures were standardized as in a training program, most of the in-class time was spent in student-led discussions on how to apply the J-modules to firm-specific problems. In this sense, our results speak more directly to the branch of the literature on management training, but can also contribute to the body of work on management consulting.

In both branches of this literature, prior work has identified similar unanswered questions (Bloom et al. (2020) for consulting and McKenzie and Woodruff (2014) for training). Future work should analyze samples that are much larger than groups of few hundred firms. It should measure the dynamics of the effects of better management for many years, instead of focusing on just the first months after the initial intervention. It should check whether management interventions generate spillovers outside the treated firms' boundaries. Moreover, it should dig

deeper into the content of management training and consulting to test what components are more effective. In recent years, two papers addressed some of these issues. Bloom et al. (2020) measured the effects of management consulting on the adoption of management practices, but not on firm performance, between seven and nine years after the intervention. Giorcelli (2019) estimated the long-term effects of management training on a large sample of Italian firms, but could not isolate vertical spillovers or complementarities within different types of training.

Our paper aimed at addressing all these open issues. It used a sample of 11,575 trained firms, at least one order of magnitude larger than most other studies in this area. It estimated the effects of training for ten years after the intervention. It measured vertical spillovers along the supply chain of trained firms and horizontal spillovers on nonapplicants. Moreover, it estimated the effect of different types of training, as well as their complementarities.

Second, our paper contributes to the literature on the economic history of WWII. On the micro side, existing works have documented the effects of WWII on female labor force participation (Goldin, 1991, Acemoglu, Autor, and Lyle, 2004; Goldin and Olivetti, 2013), the wage gaps between white and African-American workers (Margo, 1995; Collins, 2001), and the housing market (Fetter, 2016). On the macro side, research has focused on the impact of WWII on the postwar industrialization process (Fishback and Cullen, 2013; Jaworski, 2014; Koustas and Li, 2019; Garin, 2019; Bianchi and Giorcelli, 2019), the fiscal multiplier (Brunet, 2018), and the political economy of war production and government spending (Rhode, Snyder, Jr., and Strumpf, 2018). Our paper contributes to this literature by looking at the impact of WWII on the development of managerial practices that were later exported to western Europe (Giorcelli, 2019), Japan (Boel, 2003), and many other countries (Dinero, 2005).

The rest of the paper is structured as follows. Section 2 describes the origin and development of the TWI program. Section 3 describes the data. Section 4 presents the empirical framework and discusses the identification strategy. Section 5 examines the long-run effects of the TWI program on firm performance and the complementarities among different types of training. Section 6 analyzes the effects of the TWI program on nontrained firms in the economy. Section 7 investigates the influence on firm-performance of other war-related factors, as well as post-WWII turnover. Section 8 concludes.

2 Historical Background

2.1 Set-up of the TWI Program

The Training Within Industry plan was a program with the purpose of providing management training to U.S. war contractors.[1] It was established in August 1940 by the National Defense Advisory Commission after the fall of France (June 22, 1940) and was later moved to be under the jurisdiction of the newly-established War Manpower Commission (WMC) on April 18, 1942. It remained under the control of the WMC until it ceased all operations in September 1945 after Japan's surrender. Overall, the TWI maintained the same organization and functioned under the same leadership throughout its existence, in spite of the shift in jurisdiction in 1942 (Dooley, 1945; p. 106).

From the onset of WWII in September 1939, the Allied forces needed a large amount of war supplies. Many U.S. companies started receiving an increasing number of war-related orders, especially from France and Britain, that were well in excess of their productive capacity (TWI Administration, 1940). As the war escalated, it became apparent that if and when United States would join the Allies by declaring war, that event would make the situation even more critical. A great fraction of men of working age would then be called up to serve, depriving the workforce of many productive employees. The TWI program was the government's response to these concerns. It had the goal of increasing firm production and productivity to meet the increased demand. It also intended to teach U.S. firms how to train new workers and make them productive in the shortest possible amount of time.

The TWI program was set up to operate as a decentralized service. In September 1940, the TWI administration divided the U.S. into 22 geographical districts (Figure 1 and Table A.1). These districts were centered around established industrial areas, which often crossed state boundaries. Each of them had its own headquarters and was headed by a District Director.[2] To better reach peripheral areas, the TWI administrators decided to decentralize the program even further and divided each district in smaller geographical units, which we call subdistricts, headed by resident representatives. In total, they created 364 subdistricts, an average of 23 per district (Dooley, 1945; p. 7).

While the TWI program had the ambitious goal of offering management training to all U.S. war contractors, a limited budget and a lack of manpower made this initial plan not viable. More

[1] There are three main direct sources about the TWI program: the final report in December, 1945 (Dooley, 1945), the TWI bulletins (Section 1, https://play.google.com/books/reader?id=4uTNAAAAMAAJ&pg=GBS.PP7), and the progress reports (for example, https://play.google.com/books/reader?id=okxE_xNzuNgC&pg=GBS.PP1).

[2] Most District Directors were business executives who volunteered their expertise to the program. They were called "dollar-a-year" men, since they worked for free for the TWI.

importantly, the TWI administration soon realized that the success of the program hinged on the full support of trained firms' top management. Aa a result, the policy of the program became to train only firms that wanted to be part of it (TWI Administration, 1944). Accordingly, the TWI program established different application windows.[3] The only condition for applying was that firms had to be U.S. war contractors at the time of the call. In total, there were 10 application windows: one each in the years 1940, 1941, and 1945, two each in 1943 and 1944, and three in 1942. Within each subdistrict and application window, eligible firms that applied received the TWI training in the order in which they had applied.

2.2 Content of the TWI Management Training

In designing their intervention, the leaders of the TWI service adapted to the 1940s context a popular training program used during WWI.[4] The TWI training was divided into three so-called J-modules, as follows (TWI Administration, 1944):

- Job-Instructions (J-I): this module gave supervisors "practice in how to break in men on new jobs." As a result, trained firms started establishing standard procedures for operations, improving lighting, enhancing job safety measures, keeping the factory floor tidy to reduce accidents and facilitate the movement of materials, performing regular maintenance of machines, and recording the reasons for breakdowns. This module improved the factory operation practices described in Bloom et al. (2013).[5]

- Job-Relations (J-R): this module gave supervisors "practice in how to promote teamwork." Firms trained in J-R started relying more heavily on performance-based incentive systems for workers and managers, investing more in on-the-job training, experiencing a decrease in worker complaints and strikes, and defining more precisely job descriptions and responsibilities for both workers and managers. Overall, this module acted upon the practices related to human resources described in Bloom et al. (2013).

[3] Each window was closed when a target number of firms to be trained in each district was reached. However, as outlined in Section 3, even these targets were often overly optimistic and many applicant firms ended up without training.

[4] In 1917, the Emergency Fleet Corporation of the United States Shipping Board initiated a training program to increase the number of shipyard workers tenfold. To do so, they hired Charles R. Allen, a vocational instructor from Massachusetts. Allen's four-step system for training new workers—Show, Tell, Do, Check—was documented in his 1919 book *The Instructor, The Man and The Job*. This four-step methodology formed the basis for the TWI program developed over twenty years later.

[5] When this program was exported to Japan after the end of WWII, this module was split into two components: one was related to standard procedures for operations and maintenance of machines; the other one, called Job-Safety (J-S), focused on workers' safety.

- Job-Methods (J-M): this module helped supervisors "simplify and improve methods of doing a job." Firms trained in J-M learned to constantly innovate and improve their business processes. For example, they started managing more efficiently their inventory, improving production planning, and tracking production to prioritize customer orders by delivery deadline. After the end of WWII, they became more likely to develop a marketing research unit and to introduce new product lines. This module mainly operated on the inventory control and sales and order practices described in Bloom et al. (2013).

The TWI program was initially thought as a combination of management consulting and training. However, almost immediately after its implementation in 1940, the TWI administration realized that it lacked the expertise and the resources to provide effective consulting tailored to the individual needs of each war contractor. As a result, for the whole duration of the program, the TWI service focused only on writing the teaching materials associated with the three J-modules and preparing the instructors for in-plant training (Dooley, 1945; p. 16).

The content of the J-modules was constant across all trained firms and formalized in training manuals that the instructors had to follow closely. This standardization was considered a necessary step to ensure the quality of the TWI training and to allow a short-staffed organization to serve a vast group of war contractors. In spite of their rigid format, the J-modules were designed to incorporate the specific challenges faced by the employees of each trained firm. The first lecture of each J-module was a supervisor-led demonstration that had the goal to share how bad habits could improve by following the core concepts of TWI training. All the following meetings were a combination of formal lectures and "student presentations," in which the workers under training were asked to show how they could apply the TWI teachings to their job (Dooley, 1945; p. 32).

To conclude, each J-module had been designed to share at least three key characteristics. First, the content was basic, but easily digested. Instead of introducing novel concepts, the J-modules were based on accepted principles taught in an effective way.[6] As an example, upon completion of the program, trained workers were receiving an instruction card to keep in their pockets with the main takeaways of the modules they had attended (Figure A.1). Second, the program limited the time dedicated to instructors' lectures in favor of practical demonstrations in which the workers could learn by doing. This design was chosen to bring firm-specific problems into otherwise standard courses. Third, the J-modules intended to ignite a virtuous cycle of improvements within trained firms by stressing the importance of disseminating the

[6] As stated by the Director of the TWI program, "There is nothing new about TWI programs—they are built on accepted principles. The only new thing is that something was done about getting them used." (Dooley, 1945; p. ix)

TWI trainings to all workers of the firm and by teaching the scientific method to approach current and future challenges.[7]

2.3 Implementation of the Training

In addition to designing the content of the J-modules, the main responsibility of the TWI administration was the training of the instructors who would have then implemented the TWI program in the plants of applicant firms.

Most instructors were men and women with extensive business experience who were sent by their employers to volunteer for the TWI program either part-time or full-time (Dooley, 1945; p. 4). Others were paid staff already working for the Department of War or for another government agency. In spite of their different background, prospective instructors were selected based on their prior expertise in business or teaching.

Upon recruitment, candidates had to attend a "TWI institute," a formal course to become an instructor taught by a TWI staff member, either in their district of residence or in Washington, DC. Although its overall structure stayed constant, the length of a TWI institute changed from 18 hours in 1940 to 50 hours in 1944. During the first part, candidates had to attend a full J-module. The second part was divided between lectures on the fundamental concepts of a J-module and practice in teaching the course (Dooley, 1945; p. 202).[8]

Although the original plan was to prepare well-rounded instructors able to teach all three J-modules, the urgency to reach war contractors quickly induced the TWI administration to form specialists. As a result, each instructor was trained to teach only one J-module (Dooley, 1945; p. 58). After being certified, TWI instructors usually returned to their subdistrict of residence where they visited participating firms, often including their employer, in order of application to carry out their J-module of competence,

In each firm, the TWI training targeted managers, line executives, supervisors, and any other employee with "functional responsibility for planning of training" (Dooley, 1945; p. 17). In each plant of participating firms, the instructors delivered their J-modules to groups of at most 10 workers. Training one group in one module required five meetings of two hours.

Once the last group of targeted employees had attended the training, firms had to pick at least one representative who acted as a point of contact between the firm's top management and the TWI program. The TWI administration used firms' representatives to check on the

[7] To overcome a challenge, the workers were taught how to define a problem or objective, gather the facts, implement an action, and evaluate its results (Dooley, 1945; p. 41).

[8] In Section 7, we test whether the TWI program became more effective for firms trained later.

continuing adoption of TWI criteria and to collect post-training data on firm performance (Dooley, 1945; p. 166).

3 Data

We collected and digitized several types of data on the firms that applied to the TWI training. We matched separate datasets using firm name, municipality, and state. What follows is a description of these data sources and key summary statistics.

3.1 Firm-Level Government Records During WWII

We started our data collection by identifying all firms that were eligible to participate in the TWI program. Considering that the only application requirement was that firms had to be war contractors located in the United States, we retrieved the list of eligible companies from the tabulation of war supply contracts published by the Civilian Production Administration in 1946.[9] This dataset includes information on all contracts for war supplies worth at least $50,000 and awarded between June, 1940 and September, 1945. By restricting the sample to war contractors based in the United States, we obtain a list of 25,393 companies eligible to participate in the TWI program.

We then use the monthly records of the TWI Service—compiled between August 22, 1940 and September 19, 1945 and stored at the NARA Archives—to identify firms that applied to the TWI program. Out of 25,393 eligible firms, we found applications from 11,575 companies. For each application, the monthly records indicate the application date, the district and subdistrict in which the applicant firm was located, whether it eventually received the TWI training, in which of the J-modules it was trained, and the year in which each module was delivered.

Out of 11,575 applicants, 6,056 (52 percent) were eventually trained in at least one J-module. Among them, 62 percent got two J-module trainings, 24 percent got one J-module training, while the remaining firms received all three J-module interventions (Figure 2, Panel A). The bulk of the TWI training was carried out between 1943 and 1945, when the U.S. involvement in the war peaked. Only 8 percent of trained firms received a J-module intervention between 1940 and 1942, while the other firms were trained from 1943 (Figure 2, Panel B). In total, 1,750,650 workers were trained in at least one J-module (Dooley, 1945; p. 126).

From the NARA Archives, we also collected data from the plant-level surveys that the TWI administration conducted in treated firms before and after the training. Specifically, the surveys indicate whether a plant was performing each of eleven managerial practices linked to

[9] Dmitri Koustas kindly shared the digitized version of this dataset with us (Koustas and Li, 2019).

the teachings of the TWI program before the start of each J-module training, three months after the TWI training, and then each year thereafter until 1945.

Finally, we collected information on the workforce composition of applicant firms by digitizing the Selective Service's so-called replacement schedules, accessible at the NARA Archives. After the Executive Order 9279 of December 5, 1942, firms in which at least one worker was drafted had to submit a replacement schedule to their regional Bureau of Manpower Utilization. In these documents, firms described the composition of their labor force, specifically indicating the share of African-American workers and of women, as well as the average years of education and age of all their employees. Moreover, they had to propose a replacement for each drafted worker and indicate how long it would take for the new workers to become fully productive.[10] We used this information to construct the labor force composition, as well as the share of drafted employees, of each firm between 1941 and 1945.[11] Finally, firms had to indicate the name of the companies with which they had an existing contract, as well as the list of products bought from and sold to them. We used this information to identify the TWI applicants' upstream and downstream firms, as described in greater detail in Section 6.1 and Appendix C.

3.2 Data on Firm Performance

We constructed a panel dataset of firm outcomes between 1935 and 1955 using information from the annual reports collected by the Mergent Archives. Specifically, we relied on the Historical Annual Reports, a collection of over 1 million corporate documents covering more than 100 years.[12] For each firm in our sample, we searched the database of Mergent's historical annual reports available at the UCI library using firm name, municipality and state. We then digitized the information contained in the documents in order to link data on firm performance to the other data sources.

We were able to find a match for all the 11,575 applicant firms. Moreover, we retrieved information on 11,536 of the 13,857 (83 percent) nonapplicants in order to investigate the spillover effects of the program on nonparticipating firms (Section 6.2). We believe that the lower matching rate among nonapplicants is due to their smaller size. Even though there is not a formal threshold on firm size to be included in the Mergent's Historical Annual Reports Collection, in practice, publicly traded firms, firms issuing bonds, and larger firms are more likely to be included because it is easier to retrieve their financial information.

[10] Through the replacement schedules, they could also ask for draft exemptions for some categories of their workers. Managers were usually deferred "in support of national health, safety, or interest" (category II-A).

[11] While the replacement schedules started being submitted in 1942, the first schedule contained data on both 1941 and 1942.

[12] https://www.ftserussell.com/data/mergent-archives.

The annual reports are a very rich source of information. The balance sheets report, among other variables, the value of inventory, current assets, investments, and capital. The income statements separate the revenues from sales from other income sources and include information on spending in several areas, such as maintenance, repairs, worker injuries, on-the-job training. We use these variables to create measures of firm performance, such as sales revenues, productivity, and profitability, and to determine the adoption of various managerial practices. More details on the access to this data, its digitization, and the definition of the variables can be found in Appendix C.

3.3 Characteristics of Applicant Firms

In 1939, the year before the TWI program started, applicant firms had, on average, 6 plants, $191.78 million in sales (in 2020 USD), and had been in operation for eight years (Table 1, column 1). They were fairly heterogeneous in terms of employment: while the average number of employees per firm was 1,038, it ranged from a low of 543 to a high of 11,283 workers. Out of all applicant firms, 86 percent were operating in the manufacturing sector, 9 percent in transportation, 3 percent in agriculture, and 2 percent in service.[13] Between 1940 and 1945, they were awarded 0.68 war contracts per year with an average value of $25 million. Moreover, all characteristics were perfectly balanced between trained and nontrained applicants at baseline, a fact that we will explore more in Section 4 (Table 1, column 7).

At least for the manufacturing sector, it is possible to compare the applicant firms to the average U.S. firm using data from the 1939 Manufacturing Census.[14] The applicant firms included 60,521 manufacturing establishments, equal to 9.8 percent of all U.S. manufacturing establishments reported in the 1939 Manufacturing Census. Moreover, they employed 12,018,503 workers, which is equal to 40 percent of the total U.S. manufacturing labor force in 1939. In short, in spite of a relatively low share of establishments, the applicant firms made up a much higher percentage of the U.S. manufacturing workforce, denoting that they were larger than average.

In spite of their large size, applicant firms did not appear to be very productive nor profitable. Their average total factor productivity revenue (TFPR) was equal to 1.75, which is much lower than the average productivity of firms included in other studies on management interventions. For example, Bloom et al. (2013) study a sample of large Indian firms whose baseline total

[13] The information on the firm's sector is not included in the annual financial reports. To retrieve this piece of information, we imputed to each firm the 3-digit SIC codes—specifically, the classification published in 1939—associated with the firm's products listed in their first war contract.

[14] We used data from the 1939 Manufacturing Census that James Lee has digitized and kindly shared with us (Lee, 2015).

factor productivity (TFP) was 2.9, while Giorcelli (2019) focuses on small Italian firms with an average TFPR of 2.45. Moreover, in our sample, the average return on assets (ROA) was only 3 percent. In comparison, Bruhn, Dean, and Schoar (2018) analyze a sample of small Mexican firms in which the baseline ROA in the control group is 12 percent.

Of course, comparisons of TFPR and ROA across time and geography can be difficult to interpret. However, in our sample, there are more statistics that point to the fact that the applicant firms had substantial inefficiencies at baseline. For example, inventory accounted for more than 80 percent of current assets, indicating that these firms were not proficient is production planning. Interestingly, the poor management of inventory was one of the main bad practices highlighted by Bloom et al. (2013). Moreover, expenses for workers' injuries and machine repairs were both higher than spending for performance-based bonuses, suggesting that these firms were more likely to react to problems instead of setting up systems to reward positive behavior.

Finally, it should be noted that all these firms suffered major disruptions from WWII. All of them had to start producing war supplies that were in different 3-digit 1939 SIC codes.[15] Moreover, on average, 23 percent of their workers were drafted and ended up leaving their jobs.

To summarize, in spite of their large size and high sales, several pieces of evidence indicate that the TWI applicants could have greatly benefitted from management training.

4 Identification Strategy

4.1 Baseline Specification

We estimate the effect on firm performance of receiving TWI training with the following difference-in-differences event-study equation:

$$y_{i\tau} = \sum_{\tau=-5}^{10} \beta_\tau(\text{TWI}_i \cdot \text{Years after TWI}{=}\tau_{i\tau}) + \gamma \cdot \text{TWI}_i + \eta \cdot \text{App}_i + \delta_{cs\tau} + \epsilon_{i\tau}, \quad (1)$$

where the dependent variable, y_{it}, is one of several key performance metrics, such as logged sales, TFPR, and ROA of firm i in period τ.[16] TWI_i is an indicator that equals 1 if firm i received the TWI training. Years after TWI$=\tau_{i\tau}$ is a dummy equal to 1 when a calendar year

[15] Moreover, 57 percent of them had to start producing in a different 2-digit SIC code, while 37 percent had to switch to a different 1-digit SIC code.

[16] TFPR is calculated using the methodology proposed by Ackerberg, Caves, and Frazer (2015). In addition, we show that the results are robust to alternative definitions of TFPR (Figure A.3). ROA is computed as the ratio between profits and fixed gross assets. Appendix C includes a longer discussion on the construction of all variables.

is τyears before or after the year in which firm i received its first TWI training. The regression controls for the application date to the program (App_i), because it can be correlated with unobservable characteristics affecting firm performance.[17] Specifically, we include fixed effects for the application window in which firm i submitted its application to the TWI program and for the number of days between the opening of the application window and the application date. County c, sector s, and period τ fixed effects $\delta_{cs\tau}$ control for nonlinear variation in outcomes over time and within sectors and counties. Standard errors are clustered at the subdistrict and application-window level.

Unless otherwise specified, the sample contains a balanced group of firms that were always active between period -5 and period 10. Finally, because equation (1) is an event study, we need to impute values of Years after TWI=$\tau_{i\tau}$ to applicant firms that never received any training. To this end, we assume that they would have been treated contemporaneously with the first trained firm in their subdistrict and application window.[18]

4.2 Tests of the Identification Strategy

In our main specification, each coefficient β_τ captures the effect of the TWI program between period τ and -1, compared to firms operating in the same sector that applied to the program on the same date and from the same county, but which eventually did not get any training. The identifying assumption is that, conditional on nonlinear trends correlated with county-sector pairs and the application date, the performance of firms with and without TWI training would have been on parallel trends in the absence of the TWI program. While the identification assumption cannot be tested directly, here we discuss four pieces of evidence that corroborate our identification strategy.

First, as it is standard for difference-in-differences regressions, we test whether firms with and without TWI training were on parallel trends *before* the beginning of the intervention. We estimate equation (1) using only data from the five periods before the start of the TWI training. The estimated coefficients are all precisely estimated zeros. Therefore, they clearly reject the hypothesis that firms with and without TWI training experienced different pre-TWI trends in either the level or the growth rate of sales, TFPR, and ROA (Figure 3, Panels A-F). The

[17] Early applicants might have been quicker in recognizing the value of the TWI program and, therefore, might have been better managed even before the intervention.

[18] In Section 5.1, we show that estimating equation (1) on all available firms, therefore including those that entered the sample after the first period or exited before the last, leads to qualitatively similar findings. Moreover, we show that alternative choices for the timing assigned to nontrained applicants generate similar findings.

results remain unchanged if we compute the pre-TWI trends separately for any combination of J-modules, instead of pooling all trained firms in one group (Figure A.2, Panels A-F).

Second, in addition to establishing that trained and nontrained firms followed parallel pre-TWI trends, we show that firm characteristics in period -1, the year before the start of the program, do not predict the probability of receiving the training. Specifically, we regress the dummy variable TWI_i on fifteen different variables, as well as fixed effects for application date and county-sector pairs.[19] We then test the joint significance of these firm variables. The p-value associated with this test is 0.36 and, therefore fails to reject the hypothesis of no correlation between the training and firm characteristics (Table 2, Panel A, column 1). Similarly, we can replace the dependent variable with several indicators for different combinations of J-modules received by trained firms (for example, only J-R vs. J-R and J-M). Also in this case, the test of joint significance fails to find a strong correlation between different types of training and baseline firm characteristics (Table 2, Panel A, columns 2 to 8). These results are important because they indicate that the TWI administration did not choose to offer their free training to the war contractors that were deemed more likely to succeed or to the ones that could produce more war supplies, due, for example, to their larger size. Finally, we can show the same lack of correlation between the TWI training and firm-level variables, even when we split the sample by year of training (Table A.2, Panel A). These findings indicate that the TWI administration did not favor applicant firms with better characteristics even in the last stages of WWII, when the need to increase production reached its peak.

Third, we show that the baseline characteristics of the counties in which applicant firms were located do not predict the probability of receiving the training. We regress the variable TWI_i on fourteen county-level variables from the 1940 Census, imputed to each applicant firm based on its location, as well as fixed effects for application date and subdistrict-sector pairs. A p-value of 0.48 indicates that the null hypothesis of zero correlation cannot be rejected (Table 2, Panel B, column 1). Performing a similar exercise using county characteristics from the 1920 Census or the 1930 Census confirms this finding (Table 2, Panels C-D, column 1).[20] Moreover, we obtain

[19] The regressors, all measured in period -1, are: the logs of sales, value added, number of employees, number of plants, foundation year, the value of inventory, capital, current assets, investments, number of workers' strikes, monetary compensation for workers' injuries, performance-based bonus payments, number of subsidiaries, as well as distance to the nearest railroad station and distance to the nearest port. The individual coefficients are not shown, but are available upon request.

[20] The regressors are: log number of manufacturing establishments, log number of manufacturing employees, log average manufacturing wage, log total expenses in manufacturing, log of the manufacturing value added, log value of manufacturing production, farms per capita, unemployment share (available only in 1930 and 1940), log population, population per square mile, share of male residents, share of black population, share of urban population, share of illiterate population (available only in 1920 and 1930). The individual coefficients are not shown, but are available upon request.

the same result if we replace TWI_i with indicators for different combination of J-modules (Table 2, Panels B-D, columns 2-8). Finally, as shown for firm-level characteristics, we do not find any correlation between county variables and training even when we break down applicant firms by training year (Table A.2, Panels B-D). These results show that the TWI administration did not favor firms that were located in the most industrialized or economically developed areas of the United States before the start of the program. Had not this been the case, the effect of the TWI training might have conflated differential post-WWII economic trends between richer and poorer U.S. counties.

Fourth, we show that there is no autocorrelation within subdistricts between the share of firms trained in adjacent application windows (Table A.3). This finding rules out the possibility that the TWI administration favored over time applicant firms located in certain subdistricts (even though these areas were not necessarily more developed at baseline, as shown in the previous paragraph).

In short, firms with and without training (or with different types of training) were on parallel pre-TWI trends, had similar baseline characteristics, and were located in areas with similar baseline characteristics. At this point, it would be natural to wonder why, after controlling for application date and county-sector pairs, the delivery of the training was essentially as good as random. Appendix D contains several results that point to an as-good-as-random assignment of different types of TWI instructors to subdistricts and application windows.

5 Dynamic Effects of the TWI Program

In this section, we measure the dynamic effects of the TWI program over ten years after the training. In order to show more clearly that each J-module induced changes in different managerial practices, Section 5.1 limits the sample to firms that either received one J-module or no training at all. In Section 5.2, we expand the sample to all applicants, including firms that received two or three J-modules, in order to study complementarity effects among different types of management training.

5.1 Effects on Firm Performance and Managerial Practices

The estimates of equation (1) indicate that the TWI training had a long-lasting impact on firm performance. Annual sales of trained firms increased by 5.3 percent within one year of the TWI training, compared with nontrained applicants (Figure 4, Panel A). This differential effect increased to 21.7 percent in period eight and then decreased to 16 percent in period ten. The treatment effects on productivity and profitability show monotonic increases in the post-TWI

years. Specifically, the TFPR of trained firms increased by between 4.8 percent in period 1 and 36 percent in period ten, compared with nontrained firms (Figure 4, Panel B). Similarly, their ROA increased by up to 11 percent more by the tenth year after training (Figure 4, Panel C).

The results are robust to a variety of modifications to the baseline specification and sample (Figure A.3). Specifically, we replaced county-sector-period fixed effects with either firm and period or with district-sector-period fixed effects. Moreover, we clustered the standard errors at different levels of aggregation. We also imputed values of Years after TWI=$\tau_{i\tau}$ to nontrained firms using the last, instead of the first, trained firm in their subdistrict and application window. In another set of tests, we allowed for firm entry and exit, instead of using a balanced sample. Finally, we estimated TFPR using different methodologies.

There are at least two things worth noting about these results. First, since most of the training was carried out between 1942 and 1945 (Figure 2, Panel B), the war ended within three years after the program for most trained firms. However, the effects kept increasing after period three, suggesting that they were not driven by the war itself or other war-related factors, such as a higher provision of war contracts to trained firms. We will further explore this issue in Section 7. Second, the estimation of single-difference event studies indicates that firms with and without training followed the *same*, and not just parallel, trends before the TWI program (Figure 4, Panels D-F). This finding reinforces the notion that the delivery of TWI training was not correlated with preexisting conditions.

To dig deeper into the transformations that trained firms experienced after the TWI training, we can analyze the effects of the program on additional firm outcomes. The results show that trained firms sustained their growth in productivity and profitability by increasing their size. Compared with nontrained firms, the number of plants of trained firms increased by 6.4 percent by period ten, their number of employees by 13.5 percent, their number of managers by 11.4 percent, and their number of controlled subsidiaries by 5 percent (Table A.4). These findings are consistent with prior work that highlighted the positive relationship between productivity and firm size (Syverson, 2004).

While the effects of the TWI program are fairly large, their magnitude is either in line or slightly smaller than the effect sizes found in similar studies (Figure A.4). For instance, our treatment effects on sales are 4 percentage points smaller than the effects on output found by Bloom et al. (2013) and at most 6 percentage smaller than the effects on sales found by Giorcelli (2019).

In the last portion of this section, we open the black box of the TWI program by estimating the effect of different types of training. Specifically, we replace the indicator TWI_i in equation (1) with a dummy variable for each of the three J-modules. As a reminder, supervisors learned how

to break-in new employees through the J-I module, how to manage human resources through the J-R module, and how to introduce improvements to business processes through the J-M module. The results indicate that all three trainings had positive and permanent effects on sales and TFPR, while only J-R and J-M increased the ROA (Figure A.5). Moreover, the J-R module had the largest effects on every outcome and in every post-TWI period.

Next, we investigate how the J-modules led to improvements in firm performance. We start by analyzing data from the plant-level surveys the TWI administration conducted in each trained firm before the program, three months after the program, and then each year after training until 1945.[21] In general, the survey data show a clear relationship between the content of the J-modules and changes in managerial practices (Table A.5). The J-I program highlighted the importance of having the workplace properly arranged and ensuring worker safety (Dooley, 1945; p. 33). In the TWI surveys, firms that received the J-I training reported a drop in machine repairs within five years from the program and an increase in the probability of keeping a reliable record of breakdowns. Differently, firms that received the J-R module became more likely to offer on-the-job training, performance-based bonus schemes, and a process for workers to submit suggestions to higher management. Moreover, they were more likely to formalize the tasks and responsibilities attached to different positions. Finally, firms that were trained in the J-M module found new ways to cut inefficiencies or increase revenues. They reported a drop in unused inputs, as well as an increased reliance on production planning and marketing activities.

A possible caveat is that the self-reported nature of the TWI surveys may have pushed firms to overstate the adoption of good managerial practices. However, we can at least interpret these findings as evidence that trained firms were aware of the changes that each J-module was compelling them to implement. To provide additional evidence on actual changes in spending, we can validate the survey data using information from the firms' financial reports (Table 3).[22] Compared with nontrained firms, firms that received the J-I training started spending more for regular maintenance and less for machine repairs and worker injuries. Differently, firms trained in the J-R module became more likely to invest in performance-based bonus payments and on-the-job training programs. Moreover, they became less likely to experience worker strikes. Finally, firms that received the J-M training experienced a decrease in the value of their inventory, while they became more likely to increase the number of product lines and to invest in marketing activities. Remarkably, these firms did not report changes in the implementation of managerial

[21] The survey data were collected only for firms that eventually got treated. As a result, this analysis cannot include nontrained firms.

[22] Another advantage of using data from the financial reports is that we can include nontrained applicants in the analysis.

practices not related to the specific training they had received, reinforcing the link between the content of the J-modules and the firms' spending decisions.

To conclude, these findings are important for at least two reasons. First, they establish a clear connection between the TWI training and the actions implemented by trained firms. Therefore, they allow us to rule out the hypothesis that firm performance improved for factors contemporaneous to the program, but unrelated to the training itself. Second, they indicate that the adoption of good managerial practices lasted throughout the period under consideration. This fact is important to explain why we observed long-lasting effects on firm performance. As pointed out by Bloom et al. (2020), one of the main reasons why management interventions do not last is that beneficial practices are implemented in the short term, but abandoned over time. In our case, the TWI program was able to achieve a more permanent change in firm behavior, which in turn sustained a long-lasting improvement in performance. In this sense, the evidence from the TWI training is consistent with the "Toyota Way" hypothesis that better management is able to start a virtuous cycle of improvements.

5.2 Complementarity among Managerial Practices

In this section, we begin investigating why the TWI training was able to generate a long-lasting increase in firm performance and adoption of good managerial practices. At the core of the "Toyota Way" hypothesis, there is the idea that managerial practices are complementary. According to this view, by developing their managerial capital in a certain field, firms decrease the cost of improving in other areas, making it easier to become better over time. This form of complementarity is therefore able to transform a single management intervention into a stream of improvements, the so-called *kaizen* approach.

So far, testing this hypothesis has been proven to be quite challenging. In most papers that study management interventions, firms received a service—either consulting or training—that improved a "bundle" of managerial practices at the same time. These settings prevent the analysis of complementarities, because all participating firms experienced contemporaneous improvements in multiple areas. However, due to its implementation, the TWI program offers a unique opportunity to overcome these common challenges. In fact, some firms received training in only one J-module, while others received training in multiple J-modules. Therefore, we can test whether receiving training in areas x and y leads to a higher adoption of beneficial practices associated with area x, compared with receiving training in area x exclusively.

Specifically, we replace the indicator TWI_i in equation (1) with a dummy variable for each of the three J-modules. Moreover, for each J-module, we add an indicator that equals one for firms that received that module with another type of training and we interact it with a Post-$\text{TWI}_{i\tau}$

dummy. The coefficients of these new interactions measure the additional effect that stems from receiving each module together with another training, compared with receiving each module by itself. If there are no complementarities, the coefficients of these interactions will be small and not statistically significant.

The data indicate that there were moderate complementarities in the J-R and J-M modules, but none in the J-I module (Table A.6). Firms that received the J-M training with another module introduced 6.8 percent more product lines and were 2.7 percent more likely to report spending for marketing activities, compared with firms that received the J-M module alone. In a similar fashion, firms that received the J-R training with another module experienced 10.6 percent fewer strikes and were 2.6 percent more likely to report spending for on-the-job training, compared with firms that received the J-R module by itself.

Remarkably, after the TWI program was exported to Japan in the post-WWII period, the J-M training became the main inspiration for Toyota's *kaizen* approach (Dinero, 2005; p. 48). It is therefore interesting to learn that the J-M module benefitted from complementarities, one of the main features that can generate *kaizen* or, in other words, continuous improvement.

6 Effects on Other Firms

One of the peculiar features of the TWI program is the fact that it targeted a great number of large-size firms. This is at odds with most prior work on management interventions. The large scale of the program implies that we can study whether management training generated spillover effects on other nonparticipating firms.

6.1 Spillover and Selection Effects on Supply Chain

We start our analysis by examining whether the TWI training generated vertical spillovers on firms in the supply chain of trained applicants. To do so, we first identify upstream and downstream firms by combining information on their products in the replacement schedules with information on applicants' products in the war contract data.[23] Overall, we find data on 1,816 upstream and downstream firms linked to 1,572 firms that applied to the TWI program. Next, we estimate equation (1) using as dependent variable financial data from these upstream and downstream firms. Moreover, we replace the indicator TWI_i with a dummy variable for each of the three J-modules. This specification allows us to study whether the specific type of training received by an applicant had effects on the practices adopted by nontrained firms in its supply chain.

[23] More details are available in Appendix C.

The results indicate that firms in the supply chain became more productive after trained firms received the TWI training, even though their sales did not increase (Table 4). Remarkably, we find evidence that each J-module induced upstream and downstream firms to improve different managerial practices, following the same pattern that we described in Section 5.1. For example, firms in the supply chain of applicants that received the J-R training became less likely to experience worker strikes and more likely to report spending for on-the-job training. Differently, firms in the supply chain of applicants that received the J-M training decreased the size of their inventory, increased the number of product lines, and became more likely to report spending for marketing activities.

There are three additional things to note about these findings. First, these results apply to both firms that were already in the supply chain of applicants at the time of the TWI training and to firms that entered only after the TWI program. Second, with the exception of the decrease in inventory, the effect sizes tend to be smaller than the main treatment effects on applicants firms. This is to be expected because spillovers are often smaller than direct effects. Third, the J-I module does not show any evidence of vertical spillovers on the supply chain. This result could be due to its content or its smaller effects on applicant firms.

Finally, we study whether trained applicants became more likely to have better upstream and downstream firms. To do so, we focus exclusively on firms that entered the supply chain of applicants after the TWI program. Then, we regress their characteristics observed in the year before their entry in the supply chain of an applicant on the TWI_i indicator, which is one if the applicant received a form of TWI training.[24] The coefficients of TWI_i indicates that, after the TWI program, trained applicants were more likely to enter into a business relationship with more productive and larger upstream and downstream firms (Table A.7).

Overall, these findings are important because they suggest another channel through which the TWI training might have had long-lasting effects on trained firms. After fixing their own internal processes, trained applicants could sustain additional improvements by transmitting the key concepts of the TWI program to firms in their supply chain. Interestingly, the TWI program itself often emphasized the importance of spreading the J-modules within and outside the firm in order to maximize the so-called "multiplier principle" of training (Dooley, 1945; p. 6).

[24] These regressions also include fixed effects for district-sector combinations, the application window, and the number of days between the opening of the window and the firm application.

6.2 Spillover Effects on Nonapplicant Firms

The TWI program was open to all U.S. war contractors, but only 46 percent of them applied. Compared with applicants, nonapplicant firms were on average less productive, had fewer plants and employees, and reported lower sales in 1939, the year before the beginning of the TWI program (Table A.8). In spite of these differences, applicants and nonapplicants were operating in the same sectors, had a workforce with similar gender and racial composition, and were equally affected by the war. For example, comparable shares of their employees (23 percent for applicants vs. 20 percent for nonapplicants) were drafted during WWII.

In this section, we use data on nonapplicant firms to answer two related questions. First, we investigate whether the TWI training had spillover effects on the nonapplicants located near trained firms. To do so, we divide the 364 subdistricts into 52 groups of adjacent subdistricts. For each nonapplicant firm, we compute the number of applicant firms located in the same cluster of subdistricts and their average distance. Then, we regress nonapplicants' economic outcomes on one of these two measures of indirect exposure to the TWI program interacted with period dummies, in which period 0 identifies the modal treatment year within each of the 52 groups of subdistricts.[25]

The results of these regressions indicate the lack of horizontal spillovers (Figure A.6). Specifically, having a greater number of trained firms nearby does not have any correlation with worse economic outcomes. It is important to note that these findings are not necessarily ruling out the possibility that trained firms stole business from nonapplicants. Given that the U.S. economy was growing after WWII, trained firms might have gained market shares at the expense of nonapplicants even without poaching their existing customers.

Second, we investigate what the effects of the TWI program would have been, had it been implemented on a group of firms with characteristics more similar to those of nonapplicants. To do so, we replicate the estimation of equation (1) on a smaller sample of applicant firms that are matched to nonapplicants based on their features in 1939. Specifically, the matching process is based on propensity scores and uses a nearest-neighbor algorithm without replacement. The matching variables are the number of plants and employees in 1939, sales and TFPR in 1939, distance from the nearest port and railroad station, and sector. The result is a small group of 101 applicant firms with features that more closely mirror those of nonapplicants (Table A.8, column 6). Estimating the main specification on this sample leads to smaller effects of the TWI program (Figure A.7). For example, the average effect size on TFPR is 11 percent, instead of the baseline 19-percent increase.

[25] The regressions also include firm and period fixed effects.

In short, the baseline treatment effect that we estimated in Section 5.1 measures the consequences of training on the average firm that was willing to participate in the TWI program. This result, which is unbiased for willful participants, is the one that is more relevant for policy and more common in the literature on management interventions. In fact, in most cases, it is not possible to force firms to open their plants to consultants or trainers. In order to keep attrition low, especially when the sample size is small, many studies need to preselect a group of firms that showed interest in receiving the intervention. However, this last extension on nonapplicants is interesting because it might explain why these firms declined to participate. They might have expected to receive smaller benefits from the TWI training.

In addition to turning down free training during WWII, there is no evidence that the nonapplicants attempted to implement the TWI concepts during the post-WWII years, even if the program was proving effective among trained firms. This fact might be even more surprising if we consider that, after the end of the war, the TWI foundation (now a private entity) successfully exported the J-modules to 27 countries, including Japan where they inspired the implementation of lean production (Dinero, 2005; p. 42).

There are two main reasons that can explain a decreased interest for the TWI program in the United States (Dinero, 2005; p. 15). First, in the post-WWII period, U.S. firms might have perceived external trainers as an interference to their processes and a limit to their agency, instead of a source of help (Breen, 2002; p. 264). They were facing a large increase in domestic demand. Moreover, they were expanding their market share in foreign countries, due to the fact that many foreign factories had been damaged or destroyed during WWII. In this period of economic bonanza, low-productivity firms might have been shortsighted enough to consider training a luxury. This explanation is consistent with the fact that we did not find negative horizontal spillovers and with the fact that U.S. firms refocused on training only when their worldwide dominance was threatened by Japanese products in the late 1970s (Cusumano, 1985).

Second, although the TWI printed materials were available for purchase after the war, U.S. firms that wanted to receive the J-modules could not rely on the TWI administration to ensure the correct implementation of the program. The lack of carefully vetted and highly prepared TWI instructors, as well as the absence of any follow up, were possibly important enough for the success of the original program to make a self-learned TWI training much less effective.

7 War-Related Events and Other Heterogeneities

In this section, we perform several heterogeneity analyses to show how firms with different characteristics responded to the TWI program. As a general caveat, we do not leverage exoge-

nous variation along the dimensions used for these heterogeneity tests. Therefore, these results should not be interpreted as causal estimates, but as correlations between firm variables and the baseline treatment effects.

Switching to War Production So far, we have explored mechanisms that are consistent with the "Toyota-way" hypothesis. Complementarities in management practices, as well as vertical spillovers, can explain why the TWI training could generate long-lasting results. Here, we test whether the disruptions generated by WWII might have contributed to diminish the efficacy of the TWI program over time.

We start this analysis by considering the fact that many U.S. war contractors had to change their product lines to produce war items. As discussed in Section 3.3, all war contractors had to start producing war items in different 3-digit 1939 SIC codes during the course of the war, 57 percent of them in different 2-digit SIC codes, and 37 percent of them in different 1-digit SIC codes.

This change in production might have made the TWI training less effective. During the TWI in-plant training, workers might have applied the J-modules to solve the challenges they were facing while producing a given war item. After switching to a very different peacetime good, these workers might have faced a new set of problems. Although the J-modules were designed to be broad and widely applicable to different situations, it might have been difficult to apply them to different production processes.

To test this hypothesis, we estimate equation (1) separately for firms that did and did not produce war items in different 2-digit SIC codes (Table A.8). As expected, firms that had to switch to different products benefitted less from the TWI training. However, the data also indicate that changing war production was not enough to erase the effect of the program. Firms that produced items in different 2-digit SIC codes reported significant increases in sales, TFPR, and ROA.

Loss of Human Capital The draft represented another major war-related disruption to firms' processes. More than half of the male population aged 18 to 45 in 1940 (50 million people) served during WWII (Jaworski, 2014). Data from the replacement schedules indicate that all war contractors in our sample lost workers due to the draft, experiencing mobilization rates between 13 and 33 percent.

Losing experienced workers to the draft might have slowed down the application of the TWI concepts within trained firms, due to the higher influx of untrained and inexperienced

new hires. We test this hypothesis by estimating equation (1) separately for firms in different tertiles of the distribution of drafted workers (Table A.9, Panels A-C).

Firms in the highest tertile, therefore with the highest number of drafted employees, experienced a lower increase in sales, but the gap with firms in lower tertiles is not statistically significant. Differences among tertiles are even smaller for TFPR and ROA. In short, there is no evidence that the draft interfered with the efficacy of the TWI training. This result is not necessarily surprising if we think that the TWI program was introduced to teach firms how to deal with the challenges raised by the draft.

Relatedly, we can test whether the loss of human capital *after* the war was responsible for smaller treatment effects. Specifically, we use the names of the top executives that are included in the firms' financial reports to measure the share of top managers leaving the firm between the end of the training and 1955, the last year available in our sample. Between 4 and 73 percent of top executives left the company in the years after the program. Then, we estimate equation (1) separately for firms in different tertiles of the distribution of top management's turnover (Table A.9, Panels D-F).

The results on sales and TFPR indicate that the effect sizes of the program are larger for firms in which more managers stayed within the firm over the years. However, the differences among tertiles are small in magnitude and not statistically significant. For ROA, there is not a clear pattern correlated with the turnover of top managers.

These results are important because they inform the long-lasting debate about whether improvements in management belong to the firm or to individual managers. The fact that our results diminish only slightly among firms with higher turnover of managers indicates that a large part of the TWI training created firm-specific "managerial capital" (Bruhn, Karlan, and Schoar, 2010) that remained within the trained firms, even after trained managers left.These findings are at odds with Bloom et al. (2020), who document a drop in the implementation of good managerial practices when managers leave the firm; and with Huber, Lindenthal, and Waldinger (2019), who find that the loss of managers can harm a firm's profitability.

However, we think it is possible to reconcile our findings with prior work on the importance of individual managers. In the TWI context, trained firms were growing after the end of the war, significantly increasing the number of managers in their ranks (Section 5.1 and Table A.4). Therefore, these firms may have been able to retain the benefits of the TWI program by training the large number of newly hired managers on the importance of the J-modules, even if many top executives who were at the firm at the time of the TWI program were leaving. In short, had trained firms not experienced a large growth in employment, the turnover of top executives might have been a decisive factor for the long-term efficacy of the program.

Relations with the U.S. Government Here, we test whether the relationship of trained firms with the federal government strengthened after the TWI training. For example, it is plausible to assume that the Department of War might have preferred to assign war contracts to firms that had received at least one J-module, compared with applicants that did not receive any training. After all, one of the main goals of the program was to make these firms more productive in order to meet the military needs of the United States.

We test this hypothesis by estimating equation (1) with dependent variables that describe the relationship of trained firms with the federal government. We find that the number and value of war supply contracts, as well as subsidies given to war contractors after WWII, were not different between firms with and without training (Table A.9). These results show that improved outcomes are not automatically tied to trained firms having tighter economic relationships with the U.S. government.

Heterogeneities based on Sector, Location, Size, and Year of Treatment The TWI program is one of the few management interventions that targeted firms operating in different sectors, namely manufacturing, transportation, services, and agriculture. Therefore, it is interesting to study whether the benefits of the program differed across fields. The results indicate that firms operating in all sectors experienced significant increases in productivity after training, but manufacturing firms saw the largest increase (Figure A.10, Panel A). This finding is consistent with the fact the the J-modules were designed with the manufacturing sector in mind.

In the next set of results, we test whether firms in different U.S. Census Bureau regions benefitted differently from the TWI training. We find that firms in every region experienced long-lasting increases in productivity, although firms in the South had significantly lower outcomes (Figure A.10, Panel B). This finding is consistent with prior work that highlighted the slower economic development of the South in the post-WWII period (Jaworski, 2017). Moreover, it indicates that our analysis does not leverage an improvement in firm performance that is localized in one specific U.S. region, a fact that might have raised issues for our identification strategy.

Next, we test if the year in which firms received the training affected the efficacy of the program. Although firms trained early experienced more benefits, the overall pattern of treatment effects is shared by all firms that participated in the program (Figure A.10, Panel C). This finding can be explained by the fact that firms trained before 1942 received more followups from the TWI service before it shut down in 1945.

Finally, we estimate equation (1) separately for firms in different quartiles of the distribution of workforce size (Figure A.10, Panel D).[26] The results do not indicate any clear pattern that is correlated with the size of trained firms at baseline.

8 Conclusions

This paper studies the long-term effects of management training on firm performance and the adoption of managerial practices, using evidence from the Training Within Industry (TWI) Program. To perform this analysis, we linked newly digitized information on the participation of 11,575 firms to the TWI program to data from twenty years of balance sheets and income statements. Our identification strategy compares applicant firms that received the training to applicant firms that did not receive any intervention, controlling for nonlinear trends correlated with the application date and sector-county pairs. This approach is based on the fact that firms with and without training were statistically indistinguishable before the start of the training, indicating that the TWI administration did not focused its limited resources on the firms that were more likely to succeed.

We find that the TWI training had a positive and long-lasting impact on firm performance. Moreover, we establish a direct link between the content of the training and the adoption of managerial practices by trained firms. There are several channels that can explain a long-term effect of training on firm outcomes. First, we document the existence of complementarities among different types of training, which implies that receiving more training decreases the costs of improvements in new areas. Second, we find the existence of positive spillovers on firms in the supply chain of trained applicants. Third, we document that trained firms were able to retain most of the benefits of the TWI training even after a large share of their trained executives left the company. Other factors, such as a tighter business relationship with the U.S. government after training, cannot explain our results. Overall, our findings are consistent with the so-called "Toyota-way" hypothesis, which states that management interventions can generate a virtuous cycle of growth within firms.

From a historical perspective, the TWI training was a resounding success. The direct costs of the program were $155 million (in 2020 USD), which implies that the direct costs per trained supervisor were equal to $88.5.[27] Obviously, the highest costs were indirect, namely the opportunity cost of the time dedicated to the program by the employees of participating

[26] We consider period -1 to create the distribution.

[27] The direct costs were the sum of $92.3 million of appropriations and $62.5 million from the U. S. Office of Education to pay the TWI instructors (Dooley, 1945; p. 122). Moreover, 1.750 million supervisors were trained through the TWI program.

firms. Trained employees had to spend 10 hours to receive the certification in one of the three TWI modules and possibly many more hours to implement the TWI concepts within their firms. Although we do not have data on the time the trained supervisors took to spread the TWI training within their firms, the available evidence suggests that the program was cost-effective for most firms. Assuming a hourly wage for managers and supervisors of $45, supervisors would have needed to dedicate at least 3,400 hours or 21 months of their time to the TWI training in order to make the program unprofitable for their firm by the tenth year after training.[28]

We argue that these findings are important for both firms and policy makers. Firms routinely use internal training to improve the productivity of their workforce (Acemoglu and Pischke, 1998; Konigs and Vanormelingen, 2015). However, the effectiveness of these policies is usually evaluated over a limited time period, on small samples, usually without randomizing the content of the training, and without evaluating spillovers outside the trained firm (McKenzie and Woodruff, 2014). Our research addresses these open issues by showing that management training can have long-lasting effects, that the content of training matters, and that positive spillovers can take place along the supply chain. Therefore, these factors should be taken into account to ensure the success of training plans.

Are these findings applicable to today's firms? Although production processes have evolved tremendously since WWII, we think that there are several factors supporting the external validity of our results. First, the findings are relevant for more than just one industry, or for a few industries that might have disappeared or shrunk in today's economy—because our sample included over eleven thousand firms, and encompassed enterprises of different sizes with operations spanning a wide range of different industries. Second, the content of the J-modules is, perhaps surprisingly, still close to modern best practices. In fact, the managerial areas covered by the TWI training are very similar to the business principles taught in recent RCTs (see, for instance, Bloom et al., 2013). For these reasons, we believe that the findings in this paper are relevant to improving firm production in today's economy.

References

Acemoglu, Daron and Jörn-Steffen Pischke. 1998. "Why Do Firms Train? Theory and Evidence." *Quarterly Journal of Economics*, 113(1): 79–119.

[28] The hourly wage of $45 2020 USD comes from an annual wage of $5,000 1940 USD for "personnel workers" from the 1940 Occupational Outlook Handbook (https://babel.hathitrust.org/cgi/pt?id=osu.32435051428019&view=image&seq=7). Moreover, compared with nonapplicants, 6,056 trained applicant firms reported higher profits for $44.2 million by the tenth year after training. Therefore, $\frac{(\$45/\text{h} * 3{,}400\,\text{hours}) + \$88.5}{6{,}056\,\text{firms}} * 1.75\,\text{mil. trained supervisors} \simeq \$44.2\,\text{mil. higher profits}$.

Acemoglu, Daron, David H. Autor, and David Lyle. 2004. "Women, war and wages: The effect of female labor supply on the wage structure at midcentury." *Journal of Political Economy*, 112(3): 497–551.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Structural Identification of Production Functions." *Econometrica*, 83(6): 2411–2451.

Bianchi, Nicola and Michela Giorcelli. 2019. "Reconstruction Aid, Public Infrastructure, and Economic Development." Working paper, Northwestern University.

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

Bloom, Nicholas, Aprajit Mahajan, David J. McKenzie, and John Roberts. 2020. "Do Management Interventions Last? Evidence from India." *American Economic Journal: Applied Economics*, 12(2): 198–219.

Bloom, Nicholas, Benn Eifert, Aprajit Mahajan, David Mckenzie, and John Roberts. 2013. "Does Management Matter? Evidence from India." *Quarterly Journal of Economics*, 128(1): 1–51.

Boel, Bent. 2003. *The European Productivity Agency and Transatlantic Relations, 1953-61.* Copenhagen:Museum Tusculanum Press - University of Copenhagen.

Breen, William J. 2002. "Social Science and State Policy in World War II: Human Relations, Pedagogy, and Industrial Training, 1940–1945." *Business History Review*, 76(2): 233–266.

Bruhn, Miriam, Dean Karlan, and Antoinette Schoar. 2010. "What Capital is Missing in Developing Countries?" *The American Economic Review: Papers and Proceedings*, 100(2): 629–633.

Bruhn, Miriam, Karlan Dean, and Antoinette Schoar. 2018. "The Impact of Consulting Services on Small and Medium Enterprises: Evidence from a Randomized Trial in Mexico." *Journal of Political Economy*, 126(2): 635–687.

Brunet, Gillian. 2018. "Stimulus on the Home Front: The State-Level Effects of WWII Spending." Working Paper, Wesleyan University.

Cai, Jing and Adam Szeidl. 2017. "Interfirm Relationships and Business Performance." *Quarterly Journal of Economics*, 133(3): 1229–1282.

Collins, Bill. 2001. "Race, Roosevelt, and Wartime Production: Fair Employment in World War II Labor Markets." *American Economic Review*, 91(1): 272–286.

Cusumano, Michael A. 1985. *The Japanese Automobile Industry: Technology and Management at Nissan and Toyota.* Harvard University Press.

Dinero, Donald. 2005. *Training Within Industry: The Foundation of Lean.* Productivity Press; HAR/COM Edition.

Dooley, Channing R. 1945. "The Training Within Industry Report 1940-1945." War Manpower Commission. Bureau of Training.

Fetter, Daniel. 2016. "The Home Front: Rent Control and the Rapid Wartime Increase in Home Ownership." *Journal of Economic History*, 76(4).

Fishback, Price and Joseph A. Cullen. 2013. "Second World War spending and local economic activity in US counties, 1939-58." *Economic History Review*, 66(4): 975–992.

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

Garin, Andy. 2019. "Public Investment and the Spread of "Good-Paying" Manufacturing Jobs: Evidence from World War II's Big Plants." working paper.

Giorcelli, Michela. 2019. "The Long-Term Effects of Management and Technology Transfers." *American Economic Review*, 109(1): 1–33.

Goldin, Claudia. 1991. "The Role of World War II in the Rise of Women's Employment." *American Economic Review*, 81(4): 741–756.

Goldin, Claudia and Claudia Olivetti. 2013. "Shocking labor supply: A reassessment of the role of World War II on women's labor supply." *American Economic Review: Papers & Proceedings*, 103(3): 257–262.

Higuchi, Yuki, Edwin Paul Mhede, and Tetsushi Sonobe. 2016. "Short- and Longer-Run Impacts of Management Training: An Experiment in Tanzania." working paper, Nagoya City University.

Hsieh, Chang-Tai and Peter J. Klenow. 2009. "Misallocation and Manufacturing TFP in China and India." *Quarterly Journal of Economics*, 124(4): 1103–1148.

Huber, Kilian, Volker Lindenthal, and Fabian Waldinger. 2019. "Discrimination, Managers, and Firm Performance: Evidence from Aryanizations in Nazi-Germany." Working paper, University of Munich.

Ichniowski, Casey, Kathryn L. Shaw, and Giovanna Prennushi. 1997. "The Effects of Human Resource Management Practices on Productivity." *American Economic Review*, 86(3): 291–313.

Jaworski, Taylor. 2014. "'You're in the Army Now': The Impact of World War II on Women's Education, Work, and Family." *Journal of Economic History*, 74(1): 169–195.

Jaworski, Taylor. 2017. "World War II and the Industrialization of the American South." *The Journal of Economic History*, 77(4): 1048–1082.

Karlan, Dean, Ryan Knight, and Christopher Udry. 2015. "Consulting and capital experiments with microenterprise tailors in Ghana." *Journal of Economic Behavior and Organization*, 118: 281–302.

Konigs, Jozef and Stijn Vanormelingen. 2015. "The Impact of Training on Productivity and Wages: Firm-Level Evidence." *Review of Economics and Statistics*, 97(2): 485–497.

Koustas, Dmitri and Zhimin Li. 2019. "The Long-Run Effects of Government Spending on Structural Change: Evidence from Second World War Defense Contracts." *Economics Letters*, 178: 66–69.

Lee, James. 2015. "Measuring Agglomeration: Products, People, and Ideas in US Manufacturing, 1880-1990." Working paper, Harvard University.

Liker, Jeffrey K. 2004. *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer.* McGraw-Hill Education.

Margo, Robert A. 1995. "Explaining Black-White Wage Convergence, 1940-1950." *Industrial and Labor Relations Review*, 48: 470–481.

McKenzie, David J. and Christopher Woodruff. 2014. "What are we learning from business training and entrepreneurship evaluations around the developing world?" *World Bank Research Observer*, 29(1): 48–82.

Rhode, Paul W., James M. Snyder, Jr., and Koleman Strumpf. 2018. "The arsenal of democracy: Production and politics during WWII." *Journal of Public Economics*, 166: 145–161.

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

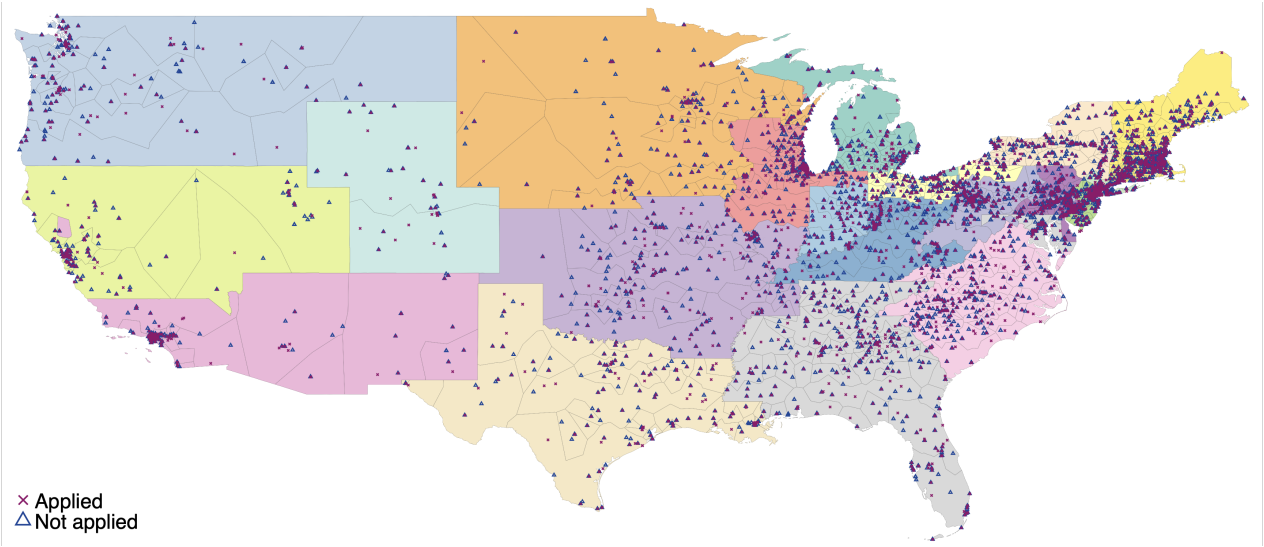
TWI Administration. 1940. "Training Workers to Meet Defense Needs." TWI Bulletin #1.

TWI Administration. 1944. "Management and Skilled Supervision (How the TWI Programs

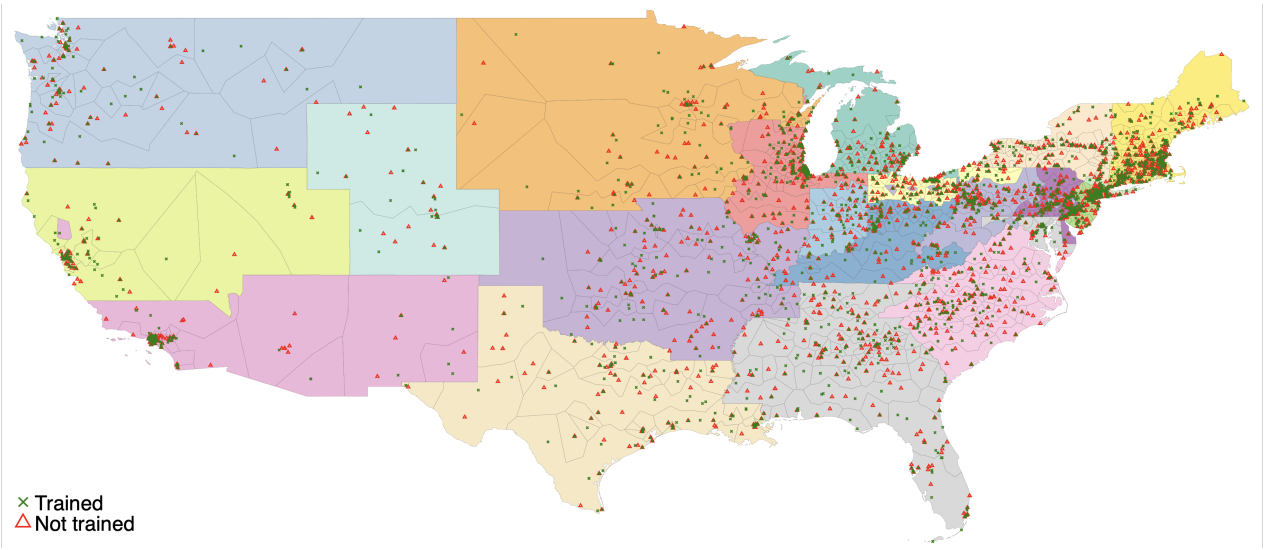
Operate)." TWI Bulletin #20.
Womack, James, Daniel Jones, and Daniel Roos. 1990. *The Machine That Changed the World.* New York:Simon and Schuster.

Figures and Tables

Figure 1: TWI Districts



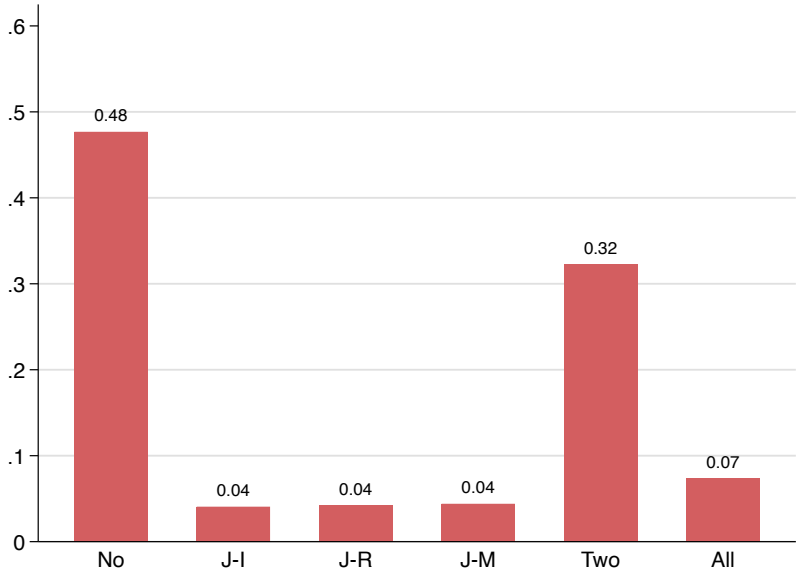
Panel A: Applicant and nonapplicant eligible firms



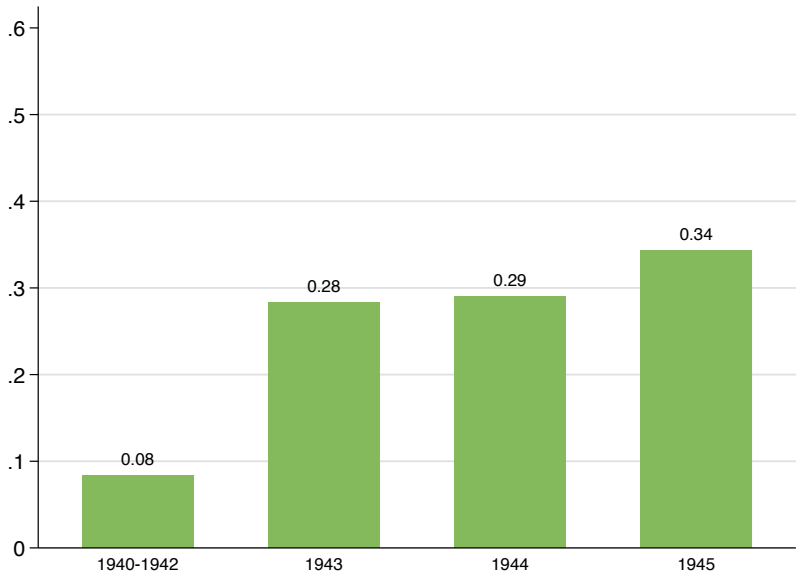
Panel B: Trained and nontrained applicant firms

Notes. Map of the 22 TWI districts. The borders within the 22 districts identify the 364 subdistricts, the level of aggregation at which the program was delivered. Panel A divides 25,393 war contractors between 11,575 firms that applied to the TWI program (purple crosses) and 13,818 firms that did not apply to the TWI program (blue triangles). Panel B divides applicant firms between 6,056 firms that received the TWI training (green crosses) and 5,519 firms that did not receive the TWI training (red triangles).

Figure 2: Distribution of TWI Training Among Applicant Firms



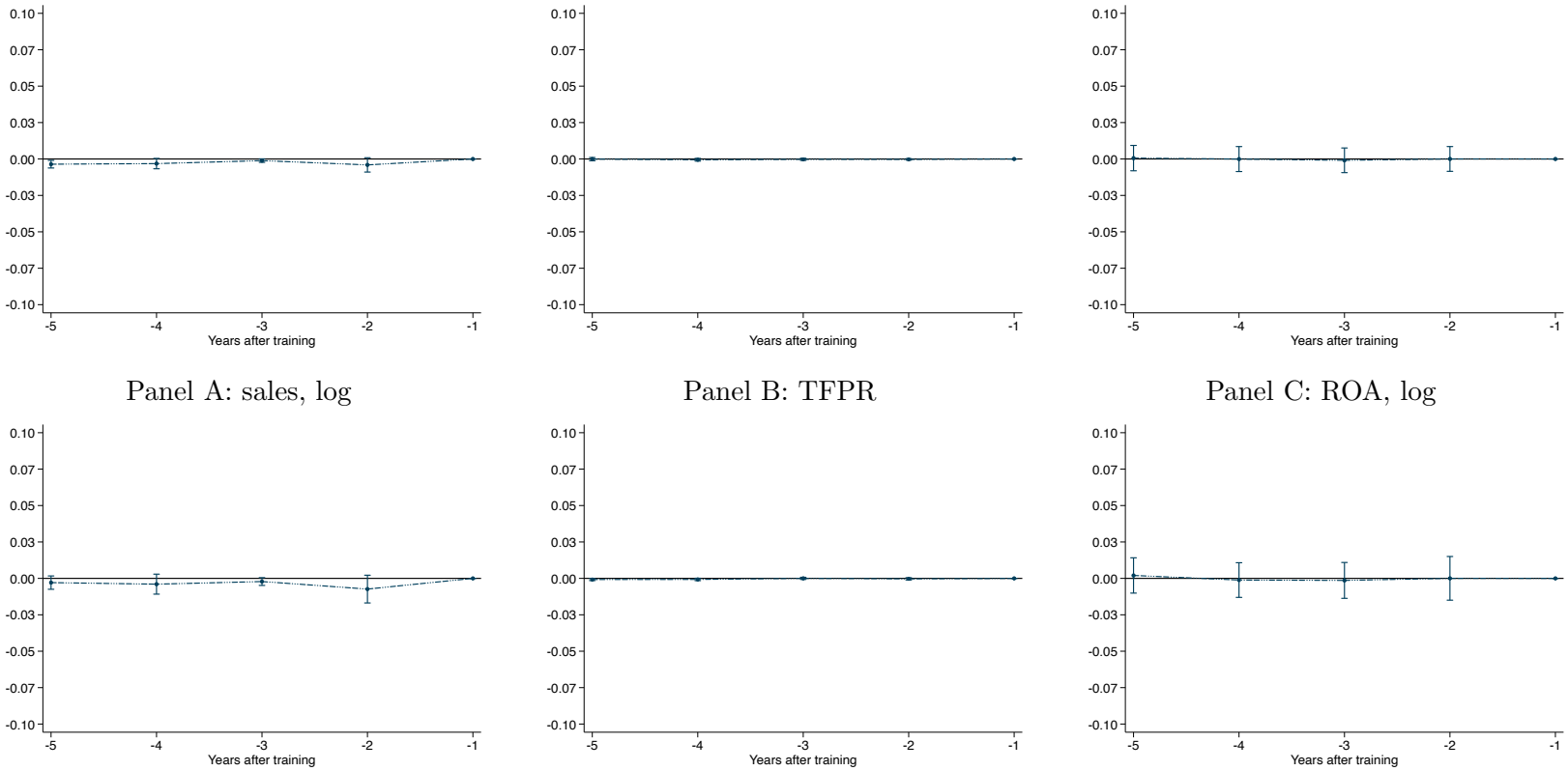
Panel A: Type of TWI training



Panel B: Year of training

Notes. This figure shows the distribution of the type of training received (Panel A) and year of training (Panel B) for the 11,575 firms that applied to the TWI program. In Panel A, *No=1* for firms that did not get any TWI intervention; J-I = 1 for firms that received the Job-Instructions training; J-R = 1 for firms that received the Job-Relations training; J-M = 1 for firms that received the Job-Methods training; Two = 1 for firms that received two trainings; All = 1 for firms that received all three trainings.

Figure 3: Pre-TWI Trends



Panel A: sales, log

Panel B: TFPR

Panel C: ROA, log

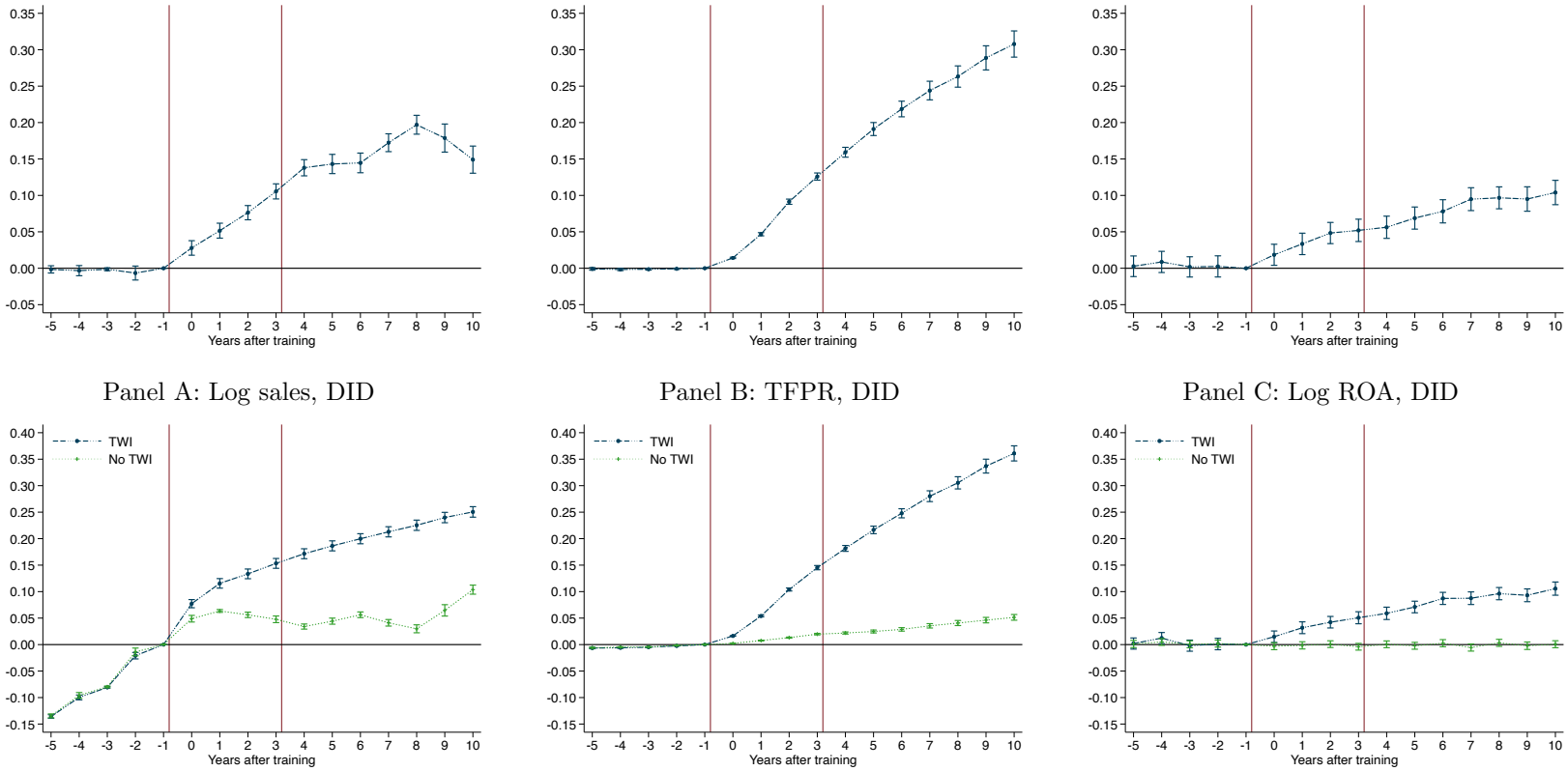
Panel D: sales, growth rate

Panel E: TFPR, growth rate

Panel F: ROA, growth rate

Notes. Coefficients of the interactions between the TWI dummy variable and pre-TWI period dummies. The omitted period is -1, one year before the TWI training. The distance from the TWI intervention for the nontrained firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. The dependent variables are: logged annual sales (Panel A), TFPR computed with the Ackerberg, Caves, and Frazer (2015) method (Panel B), logged return on assets (ROA), computed as the ratio between profits and fixed gross assets (Panel C), annual growth rate of sales (Panel D), annual growth rate of TFPR (Panel E), and annual growth rate of ROA (Panel F). The sample used for these graphs includes all 11,575 firms that applied to the TWI program. The vertical bars denote 95 percent confidence intervals. The standard errors are clustered at the level of subdistricts and application windows.

Figure 4: Effects of TWI Training on TFPR



Panel A: Log sales, DID

Panel B: TFPR, DID

Panel C: Log ROA, DID

Panel D: Log sales, Single Difference

Panel E: TFPR, Single Difference

Panel F: Log ROA, Single Difference

Notes. Panels A-C show difference-in-differences coefficients, while panels D-F show single-difference coefficients. These regressions also include fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The distance from the TWI intervention for the nontreated firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. The dependent variables are log sales (Panels A and D), TFPR computed with the Ackerberg, Caves, and Frazer (2015) method (Panels B and D), log of return on assets (Panels C and F). The sample includes applicant firms that either received only one TWI training or no training at all. The vertical bars denote 95 percent confidence intervals. The first vertical red line identifies the beginning of the TWI program. The second vertical red line identifies the end of World War II for most applicant firms. The standard errors are clustered at the level of subdistricts and application windows.

Table 1: Summary Statistics for 11,575 Applicants to the TWI Program

	All Applicant Firms				Trained Firms	Nontrained Firms	Difference
	Mean	St. Dev.	Min	Max	Mean	Mean	*p*-value
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Data from annual financial statements in 1939							
Plants	6.08	1.90	3	14	6.09	6.07	0.588
Employees	1,038.32	378.53	543	11,283	1,036.99	1,039.77	0.686
Foundation year	1931	3.73	1912	1936	1930.96	1931.05	0.229
Agriculture	0.03	0.16	0	1	0.03	0.03	0.678
Manufacturing	0.86	0.34	0	1	0.86	0.86	0.746
Transportation	0.09	0.29	0	1	0.09	0.09	0.400
Services	0.02	0.14	0	1	0.02	0.02	0.624
Sales	191.78	77.12	75.74	2,506.72	191.27	192.34	0.444
Current assets	18.40	8.02	5.31	170.72	18.47	18.33	0.341
Total assets	64.45	25.79	23.03	900.89	64.40	64.51	0.808
TFPR	1.75	0.61	0.33	3.80	1.74	1.75	0.664
ROA	0.03	0.01	0.01	0.09	0.03	0.03	0.010
Inventory	15.37	6.24	5.21	181.02	15.36	15.38	0.854
Injuries	10.40	5.09	2.40	139.78	10.33	10.47	0.141
Repairs	15.37	7.26	4.04	226.42	15.28	15.48	0.134
Bonus payments	8.27	3.97	1.97	118.42	8.22	8.32	0.176
Panel B: Workforce data from replacement lists in 1941							
Share African-Americans	0.06	0.01	0.01	0.12	0.06	0.06	0.180
Share women	0.05	0.01	0.01	0.09	0.05	0.05	0.703
Years of education	8.51	1.36	5	12	8.50	8.52	0.457
Age of workforce	28.98	3.53	23	40	28.96	29.00	0.611
Panel C: WWII-related data from replacement lists and war contracts							
Share drafted employees (1942-1945)	0.23	0.03	0.13	0.33	0.23	0.23	0.180
Switched 3-digit SIC (1940-1945)	1	0	1	1	1	1	—
Switched 2-digit SIC (1940-1945)	0.57	0.50	0	1	0.57	0.57	0.556
Switched 1-digit SIC (1940-1945)	0.37	0.48	0	1	0.37	0.38	0.257
Number of contracts (1940-1945)	0.68	2.61	0	152	0.68	0.68	0.719
Value of contracts (1940-1945)	25.11	293.01	0	27,145.94	23.87	26.48	0.507

Notes. Summary statistics for 11,575 firms that applied to the TWI program. Column 7 shows the p-value of the difference between columns 5 and 6. Panel A includes variables from the firm's financial statements in 1939. *Plants* is the total number of plants per firm; *Employees* is the number of employees per firm; *Agriculture*, *Manufacturing*, *Transportation*, and *Services* are indicators that equal one if a firm operates in agriculture, manufacturing, transportation, or services, respectively; *Sales, Current assets, Total assets, Inventory, Injuries, Repairs, Bonus payments* are expressed in 2020 million USD; *TFPR* is the log of total factor productivity revenue, estimated using the method proposed by Ackerberg, Caves, and Frazer (2015); *ROA* is returns on assets, measured as profits over total assets. Panel B includes variables from the wartime replacement schedules in 1941. *Share African-Americans* is the share of African-American workers; *Share women* is the share of female workers; *Year of education* is the average years of formal education of the workforce; and *Age of workforce* is the average age of the workers. Panel C shows data on the firm exposure to WWII from either the replacement schedules and the war contracts. *Share drafted employees* is the share of employees drafted between 1942 and 1945; *Switched x-digit SIC* equals one if a firm's war products were in different x-digit 1939 SIC codes; *Number of contracts* is the yearly number of war contracts between 1940 and 1945; and *Value of contracts* is the yearly value of war contracts (in 2020 million USD).

Table 2: Correlation between Firm and County Characteristics and Training Received

	TWI	J-I	J-R	J-M	J-I and J-R	J-R and J-M	J-I and J-M	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Regressions of training variables on firm characteristics in period -1								
p-value of joint significance	0.36	0.85	0.70	0.54	0.92	0.50	0.77	0.66
Observations	11,571	11,571	11,571	11,571	11,571	11,571	11,571	11,571
Panel B: Regressions of training variables on county characteristics in year 1940								
p-value of joint significance	0.48	0.25	0.36	0.21	0.45	0.47	0.86	0.51
Observations	11,034	11,034	11,034	11,034	11,034	11,034	11,034	11,034
Panel C: Regressions of training variables on county characteristics in year 1930								
p-value of joint significance	0.08	0.44	0.26	0.34	0.07	0.14	0.69	0.66
Observations	11,148	11,148	11,148	11,148	11,148	11,148	11,148	11,148
Panel D: Regressions of training variables on county characteristics in year 1920								
p-value of joint significance	0.45	0.32	0.32	0.35	0.15	0.46	0.68	0.50
Observations	11,160	11,160	11,160	11,160	11,160	11,160	11,160	11,160

Notes. Panel A shows the p-value of the test of joint significance of the coefficients of fifteen firm characteristics observed in period -1. The variables are: the logs of sales, value added, number of employees, number of plants, foundation year, the value of inventory, capital, current assets, investments, number of workers' strikes, monetary compensation for workers' injuries, performance-based bonus payments, number of subsidiaries, as well as distance to the nearest railroad station and distance to the nearest port. The regressions also include fixed effects for county-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Panels B, C, and D show the p-value of the test of joint significance of the coefficients of several county characteristics measured in 1940, 1930, and 1920, respectively. The county-level variables are imputed to firms based on their location. These regressions include: log population, log of the manufacturing value added, log number of manufacturing establishments, log number of manufacturing employees, log average manufacturing wage, log total expenses in manufacturing, log value of manufacturing production, farms per capita, unemployment share (available only in 1930 and 1940), population per square mile, share of male residents, share of black population, share of urban population, share of illiterate population (available only in 1920 and 1930). County data are from IPUMS NHGIS, www.nhgis.org. The regressions also include fixed effects for subdistrict-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Standard errors are clustered at the subdistrict-application window level.

Table 3: Adoption of Managerial Practices

	Log repairs (1)	Log maintenance (2)	Log injuries (3)	Log bonus (4)	Log strikes (5)	Prob training (6)	Log inventory (7)	Log product lines (8)	Prob marketing (9)
J-I x Period 1	-0.010***	0.005***	-0.020**	-0.000	0.000	-0.001	-0.010	0.002	-0.001
	(0.000)	(0.000)	(0.010)	(0.000)	(0.001)	(0.001)	(0.011)	(0.003)	(0.001)
J-I x Period 5	-0.033***	0.014***	-0.043***	-0.000	0.006***	-0.051***	0.007	-0.002	-0.028***
	(0.001)	(0.001)	(0.010)	(0.000)	(0.002)	(0.006)	(0.010)	(0.004)	(0.005)
J-I x Period 10	-0.057***	0.025***	-0.064***	0.000	0.030***	-0.062***	0.008	-0.004	-0.064***
	(0.001)	(0.001)	(0.011)	(0.000)	(0.004)	(0.006)	(0.012)	(0.005)	(0.007)
J-R x Period 1	-0.000	0.000	0.004	0.414***	-0.000	0.043***	-0.005	0.002	0.001*
	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.010)	(0.011)	(0.002)	(0.001)
J-R x Period 5	-0.000	0.000	0.011	0.431***	-0.268***	0.628***	0.008	-0.001	-0.023***
	(0.000)	(0.000)	(0.010)	(0.000)	(0.006)	(0.025)	(0.010)	(0.004)	(0.004)
J-R x Period 10	-0.000	-0.000	-0.000	0.459***	-0.268***	0.927***	0.033***	-0.000	-0.061***
	(0.000)	(0.000)	(0.009)	(0.001)	(0.006)	(0.008)	(0.010)	(0.005)	(0.006)
J-M x Period 1	0.000	0.000	-0.001	0.000	0.000	0.001	-0.015***	0.015***	0.043***
	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	(0.001)	(0.005)	(0.004)	(0.010)
J-M x Period 5	-0.000	0.000	0.001	-0.000	0.007***	-0.038***	-0.071***	0.538***	0.661***
	(0.000)	(0.000)	(0.009)	(0.000)	(0.002)	(0.005)	(0.006)	(0.009)	(0.024)
J-M x Period 10	-0.000	0.000	0.006	-0.000	0.013***	-0.062***	-0.100***	0.606***	0.930***
	(0.000)	(0.000)	(0.009)	(0.000)	(0.002)	(0.006)	(0.007)	(0.009)	(0.008)
Observations	67,472	67,472	67,472	67,472	67,472	67,472	67,472	67,472	67,472
R^2	0.142	0.131	0.141	0.170	0.165	0.559	0.406	0.445	0.648

Notes: This table shows the coefficients of the interactions between the training variables and three selected period dummies (out of 15 period dummies included). The omitted period is the year before the TWI training (period -1). The distance from the TWI intervention for the nontreated firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. All the dependent variables are logged with the exception of two dummy variables that measure whether firms reported expenditures for on-the-job training (column 6; mean=0.06) or for marketing activities (column 9; mean=0.05). The regressions also include the treatment variables in isolation, as well as fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The sample includes applicant firms that either received only one TWI training or no training at all. Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Effects on Upstream and Downstream Firms

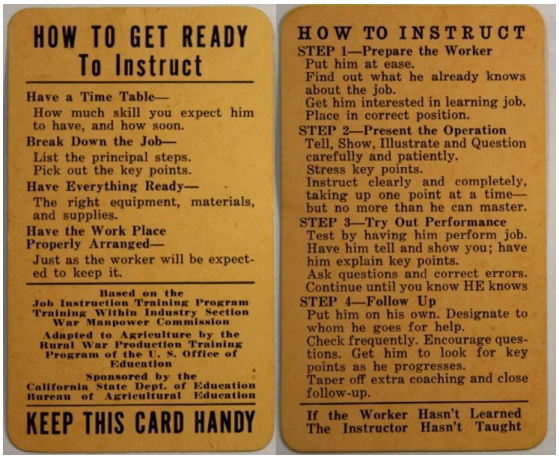
	Firms in supply chain before TWI			Firms entering supply chain after TWI		
	J-I x Post	J-R x Post	J-M x Post	J-I x Post	J-R x Post	J-M x Post
	(1)	(2)	(3)	(4)	(5)	(6)
Log sales	-0.001	0.009	0.005	-0.003	-0.001	-0.001
	(0.011)	(0.012)	(0.012)	(0.004)	(0.004)	(0.003)
TFPR	0.076***	0.055**	0.060***	0.032**	0.052***	0.058***
	(0.027)	(0.025)	(0.016)	(0.014)	(0.015)	(0.012)
Log repairs	0.048	0.033	0.026	-0.014	-0.006	0.003
	(0.065)	(0.056)	(0.046)	(0.042)	(0.038)	(0.039)
Log maintenance	-0.017	-0.134**	0.025	-0.021	-0.039	-0.038
	(0.074)	(0.053)	(0.058)	(0.108)	(0.083)	(0.110)
Log injuries	0.050	0.011	0.018	-0.031	-0.081	-0.025
	(0.066)	(0.053)	(0.052)	(0.072)	(0.066)	(0.052)
Log bonus	0.018	0.046	0.089	0.036	0.021	-0.002
	(0.035)	(0.039)	(0.054)	(0.060)	(0.057)	(0.053)
Log strikes	0.026	-0.143**	0.034	-0.006	-0.127**	-0.019
	(0.059)	(0.063)	(0.037)	(0.038)	(0.053)	(0.050)
Prob training	-0.052*	0.431***	-0.035*	-0.040	0.536***	-0.029
	(0.027)	(0.082)	(0.020)	(0.052)	(0.062)	(0.036)
Log inventory	-0.025	-0.022	-0.413***	-0.080**	-0.027	-0.523***
	(0.044)	(0.021)	(0.093)	(0.032)	(0.038)	(0.110)
Log product lines	-0.013	0.027	-0.004	-0.037	-0.010	0.069***
	(0.035)	(0.025)	(0.024)	(0.030)	(0.034)	(0.022)
Prob marketing	-0.023	-0.038*	0.346***	0.011	-0.008	0.464***
	(0.017)	(0.019)	(0.057)	(0.02)	(0.013)	(0.048)

Notes. The sample used for this table includes 1,816 upstream and downstream firms linked to 1,572 firms that applied to the TWI program. Each row shows the coefficients of the interactions between the training variables and period dummies from a different regression. In these difference-in-differences specifications, the treatment effects measure the difference between firms linked to applicants that received a certain form of management training and firms linked to applicants that did not receive any training, and between each year and the year just before the implementation of the TWI program (period -1). These regressions also include the training variables in isolation, as well as fixed effects for county-sector-period combinations, the application window of the applicant firm, and the number of days between the opening of the window and the firm application. The distance from the TWI intervention is imputed using the distance from the TWI program of the applicant firm with which the upstream/downstream firm has a contractual relationship. If an upstream/downstream firm joins the supply chain of an applicant firm after the latter received the TWI program, period 0 identifies the year of entry in the supply chain. The training variables are: J-I = 1 for firms linked to applicants that received the Job-Instructions training; J-R = 1 for firms linked to applicants that received the Job-Relations training; J-M = 1 for firms linked to applicants that received the Job-Methods training. Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

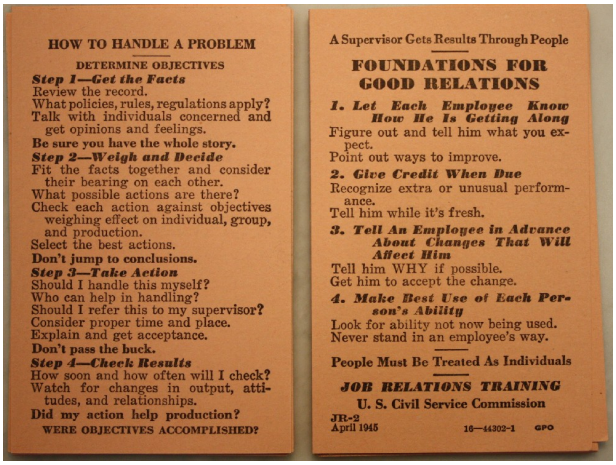
Online Appendix

A Additional Figures and Tables

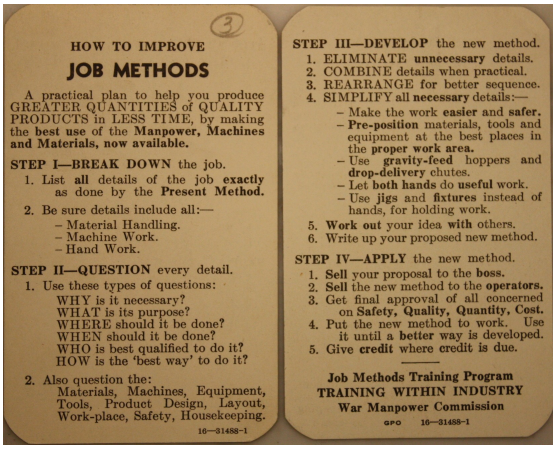
Figure A.1: Instruction Cards for Each J-Module



Panel A: Job Instruction



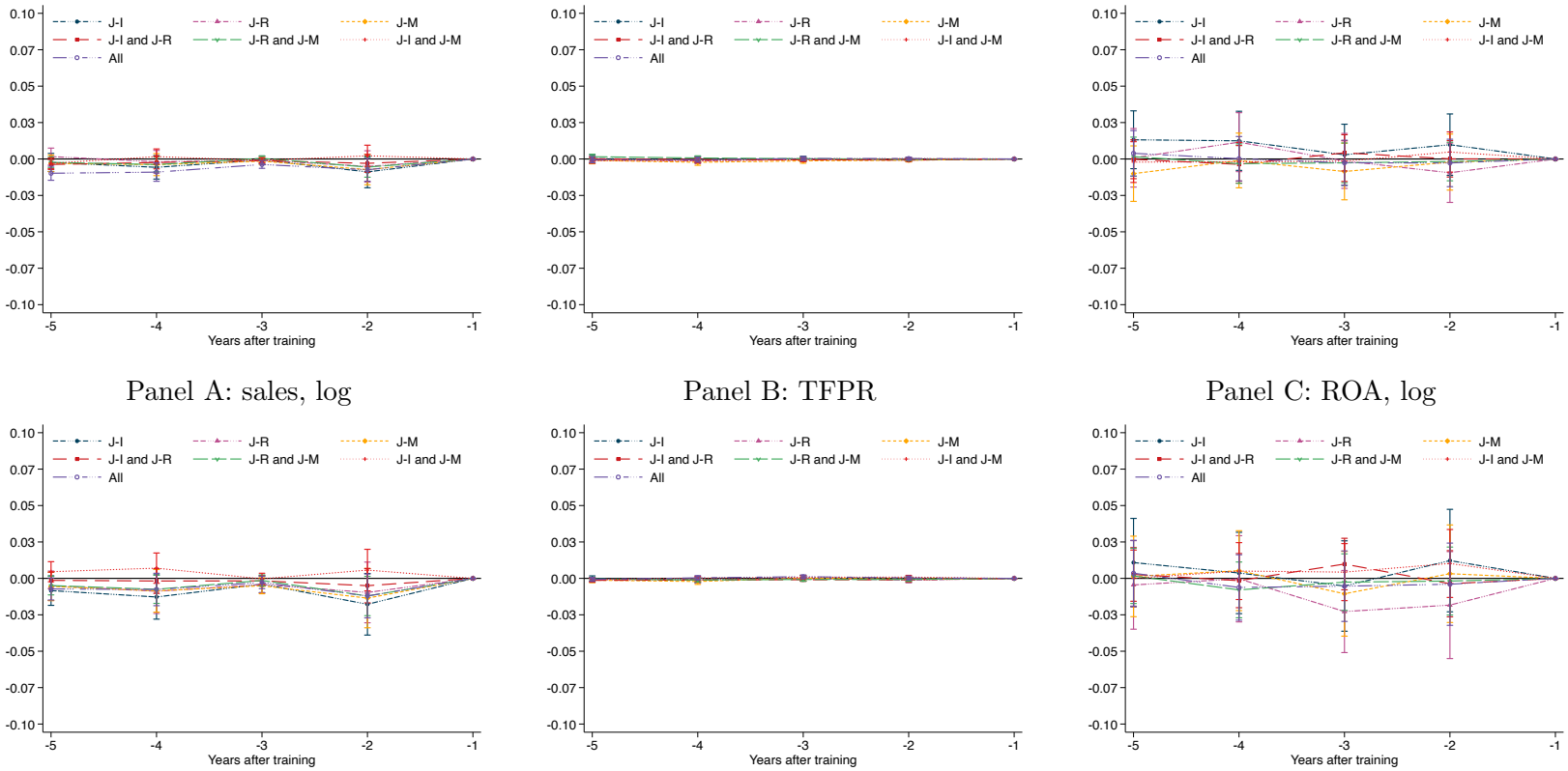
Panel B: Job Relations



Panel C: Job Methods

Notes. The TWI training was designed to be simple and immediately usable. To emphasize these features, the TWI administration prepared instruction cards for each J-Module that trained supervisors could keep in their pockets while they were working. These instruction cards reported the main 4 steps (and their sub-steps) that characterized each module. It is important to note that these 4 "theoretical" steps did not constitute the entirety of the TWI trainings. A large portion of the class time was dedicated to in-class student presentations on how to apply TWI concepts in their firms.

Figure A.2: Pre-TWI Trends For Different Types of TWI Training



Panel A: sales, log

Panel B: TFPR

Panel C: ROA, log

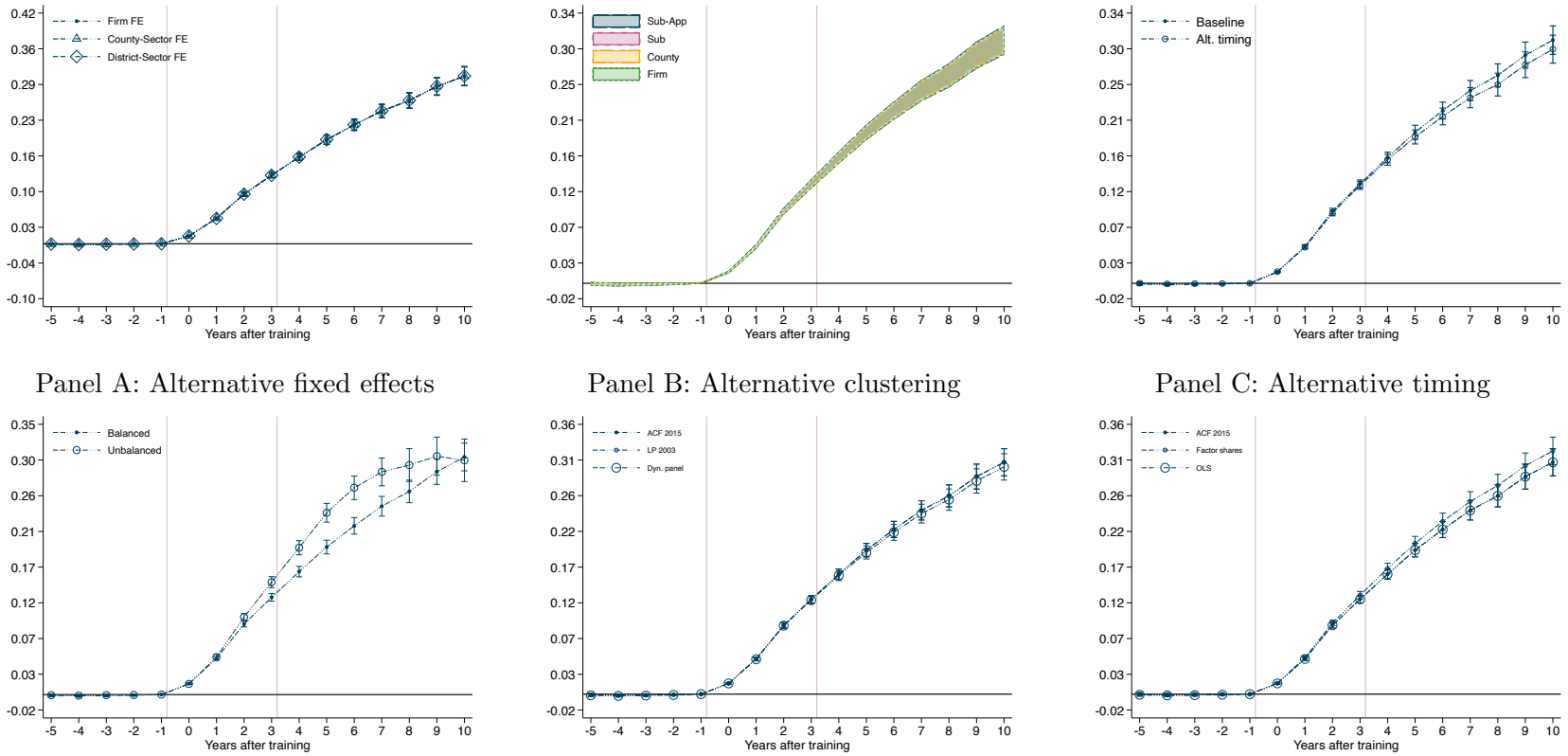
Panel D: sales, growth rate

Panel E: TFPR, growth rate

Panel F: ROA, growth rate

Notes. Coefficients of the interactions between all the different training variables and pre-TWI period dummies. The omitted period is -1, one year before the TWI training. The distance from the TWI intervention for the nontrained firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. The dependent variables are: logged annual sales (Panel A), TFPR computed with the Ackerberg, Caves, and Frazer (2015) method (Panel B), logged return on assets (ROA), computed as the ratio between profits and fixed gross assets (Panel C), annual growth rate of sales (Panel D), annual growth rate of TFPR (Panel E), and annual growth rate of ROA (Panel F). The sample used for these graphs includes all 11,575 firms that applied to the TWI program. The vertical bars denote 95 percent confidence intervals. The standard errors are clustered at the level of subdistricts and application windows.

Figure A.3: Robustness Checks on the Effects on TFPR



Panel A: Alternative fixed effects

Panel B: Alternative clustering

Panel C: Alternative timing

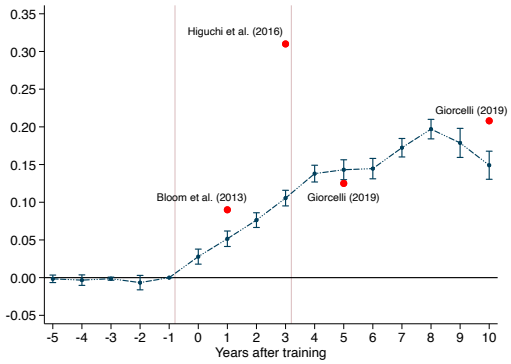
Panel D: Unbalanced sample

Panel E: Alternative TFPR Estimation (1)

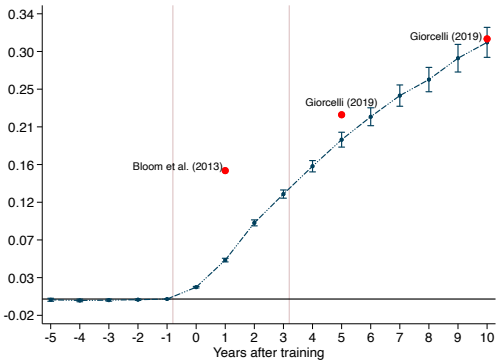
Panel F: Alternative TFPR Estimation (1)

Notes. In all panels, the dependent variable is the TFPR, estimated using the Ackerberg, Caves, and Frazer (2015) method. Panel A shows treatment effects including firm fixed effects or district-sector-time fixed effects, instead of county-sector-time fixed effects. Panel B shows the effect of clustering the standard errors at the subdistrict, county, or firm level, instead of at the level of subdistricts and application windows. In panel C, the distance from the TWI intervention for the nontrained firms is imputed using the distance from the TWI intervention of the last, instead of the first, trained firm in the same subdistrict and application window. Panel D shows the the effect of using an unbalanced sample, allowing for firm exit and entry. Panels E and D use alternative methodologies to compute TFPR.

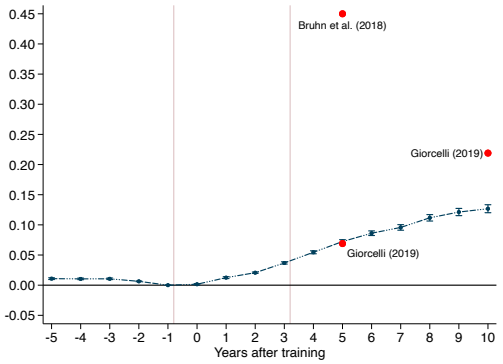
Figure A.4: Comparison with Effect Sizes in the Literature



Panel A: Log sales



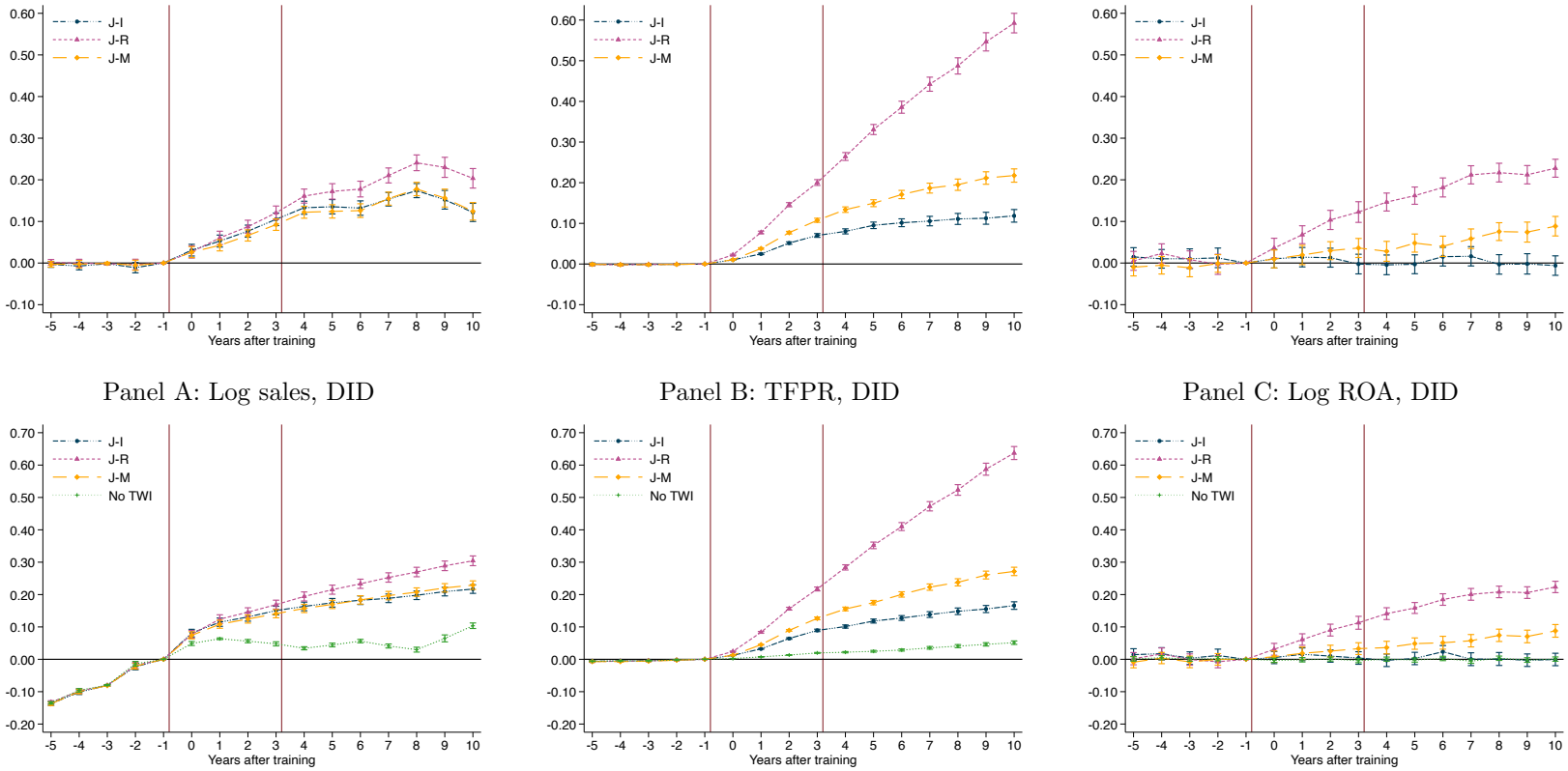
Panel B: TFPR



Panel C: Employees

Notes. Bloom et al. (2013) show results on output, instead of sales, and TFP, instead of TFPR.

Figure A.5: Effects of Different J-Modules on Firm Performance



Panel A: Log sales, DID

Panel B: TFPR, DID

Panel C: Log ROA, DID

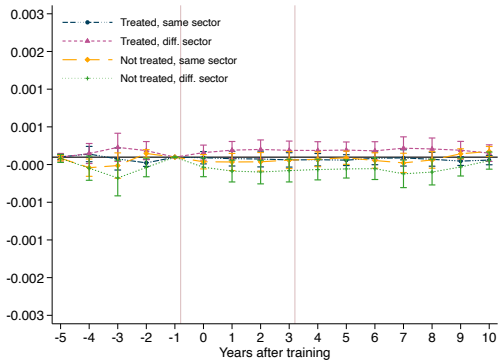
Panel D: Log sales, Single Difference

Panel E: TFPR, Single Difference

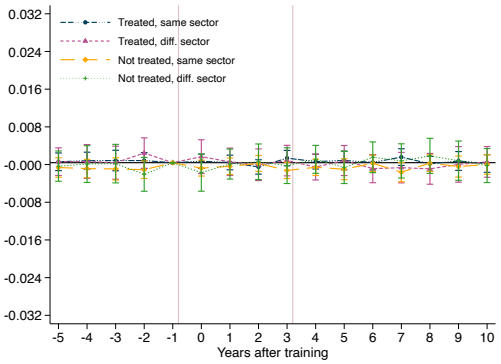
Panel F: Log ROA, Single Difference

Notes. Panels A-C show difference-in-differences coefficients, while panels D-F show single-difference coefficients. These regressions also include fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The distance from the TWI intervention for the nontreated firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. The dependent variables are log sales (Panels A and D), TFPR computed with the Ackerberg, Caves, and Frazer (2015) method (Panels B and D), log of return on assets (Panels C and F). The sample includes applicant firms that either received only one TWI training or no training at all. The vertical bars denote 95 percent confidence intervals. The first vertical red line identifies the beginning of the TWI program. The second vertical red line identifies the end of World War II for most applicant firms. The standard errors are clustered at the level of subdistricts and application windows.

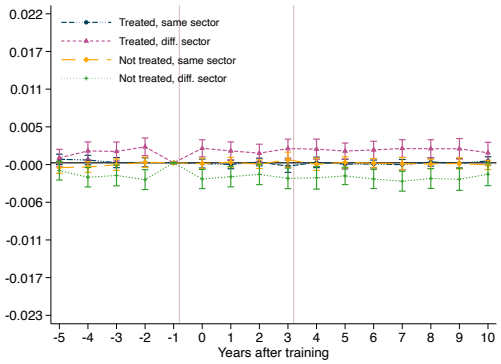
Figure A.6: Horizontal Spillovers



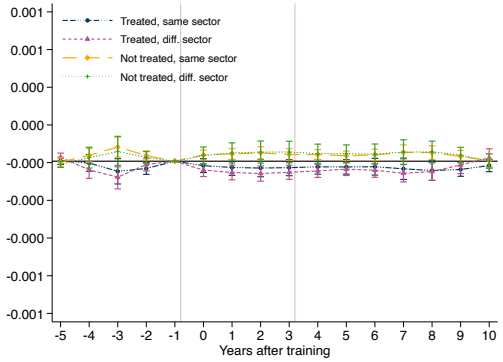
Panel A: Sales, # nearby applicants



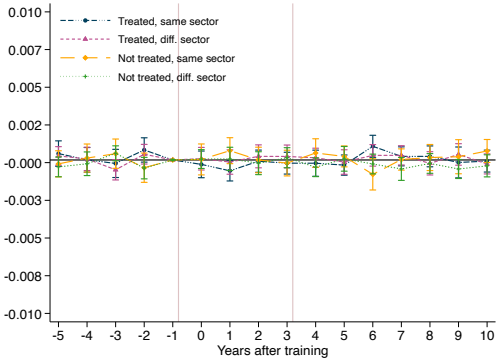
Panel B: TFPR, # nearby applicants



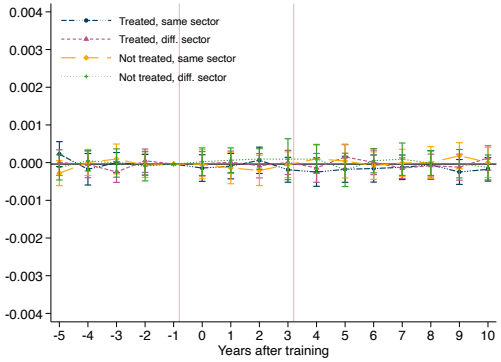
Panel C: ROA, # nearby applicants



Panel D: Sales, distance to applicants



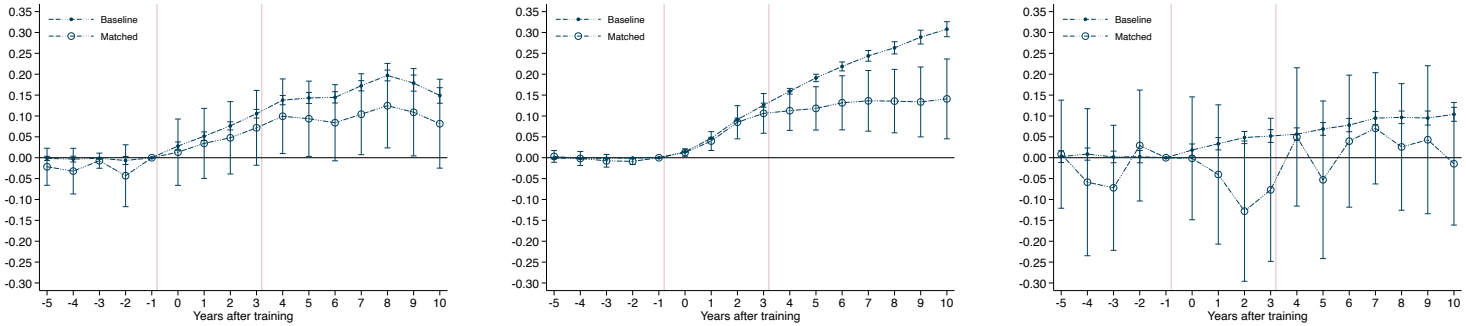
Panel E: TFPR, distance to applicants



Panel F: ROA, distance to applicants

Notes. The sample used for these graphs includes 11,536 war contractors that did not apply to the TWI program and for which we found financial statements (out of 13,818 total nonapplicants). In panels A-C, the coefficients shown are the interactions between the log number of applicants in the area and period dummies. The number of applicants is divided between trained and nontrained firms and between firms in the same or different sector (agriculture, manufacturing, transportation, or services). Panels D-F use the log of the average distance from applicant firms, instead of the log number of firms. In order to compute the number of nearby applicant firms and their average distance to nonapplicant firms, we divided the 364 subdistricts into 52 groups of 7 adjacent subdistricts. Period 0 coincides with the modal treatment year within each of the 52 groups of subdistricts. All regressions also include firm fixed effects. The standard errors are clustered at the subdistrict level.

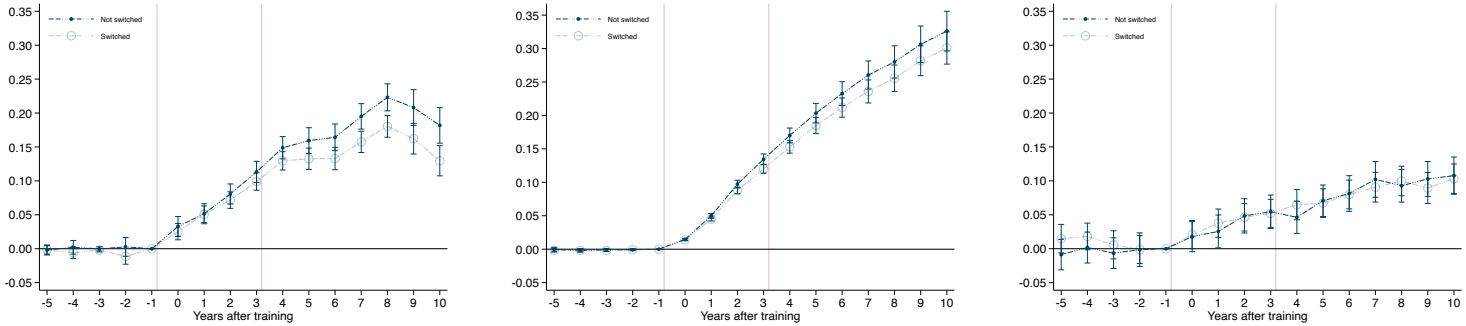
Figure A.7: Effects of TWI Training on Matched Sample



Panel A: Log sales Panel B: TFPR Panel C: Log ROA

Notes. These graphs replicate the estimation shown in Figure 4 on a smaller sample of applicant firms matched to nonapplicants based on their characteristics in 1939. The matching was based on propensity scores, using a nearest-neighbor algorithm without replacement. The variables used for the matching were the number of plants and employees in 1939, sales and TFPR in 1939, distance from the nearest port and railroad station, and sector. The standard errors are clustered at the level of subdistricts and application windows.

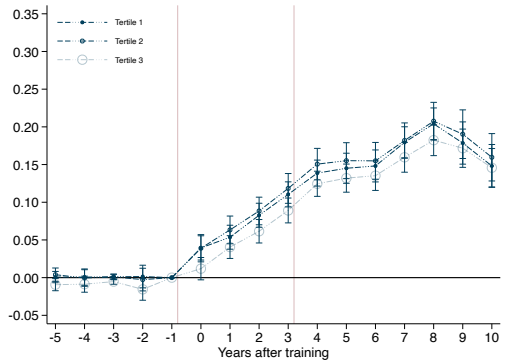
Figure A.8: Change in Status Quo, Switching to War Production



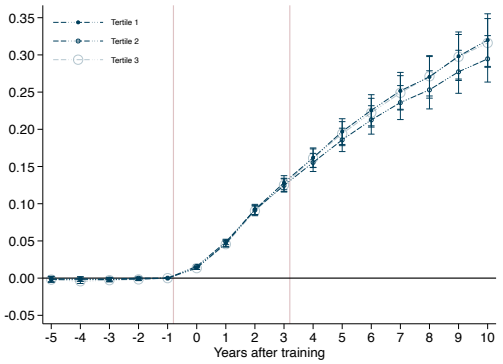
Panel A: Log sales Panel B: TFPR Panel C: Log ROA

Notes. These graphs show the interactions between the training variables and the period dummies distinguishing between firms that did or did not switch production during World War II. Specifically, *Switched 2-digit SIC* equals one if a firm's war products were in different 2-digit 1939 SIC codes. The standard errors are clustered at the level of subdistricts and application windows.

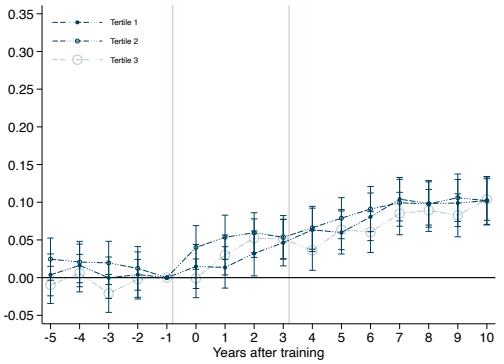
Figure A.9: Loss of Human Capital



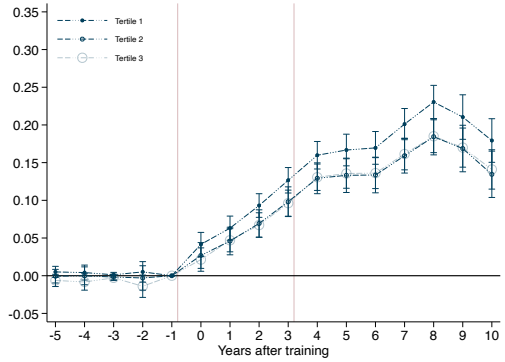
Panel A: Sales, share of drafted workers



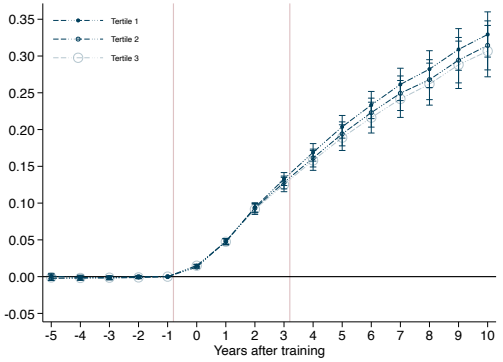
Panel B: TFPR, share of drafted workers



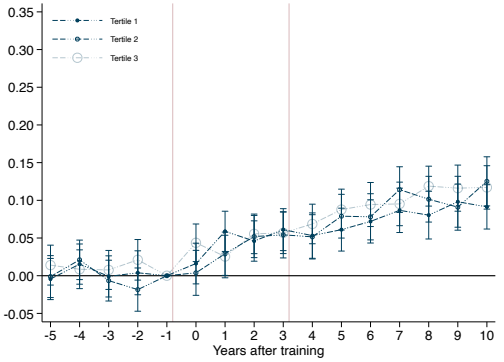
Panel C: ROA, share of drafted workers



Panel D: Sales, share of managers leaving



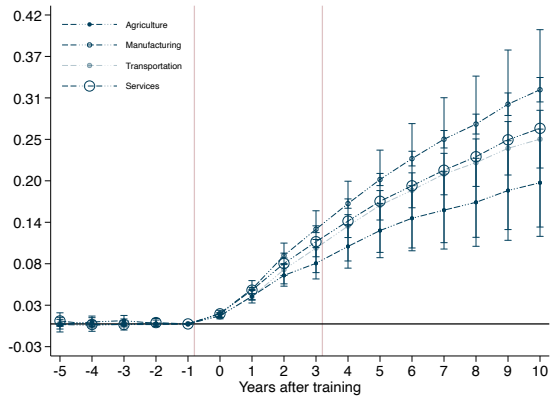
Panel E: TFPR, share of managers leaving



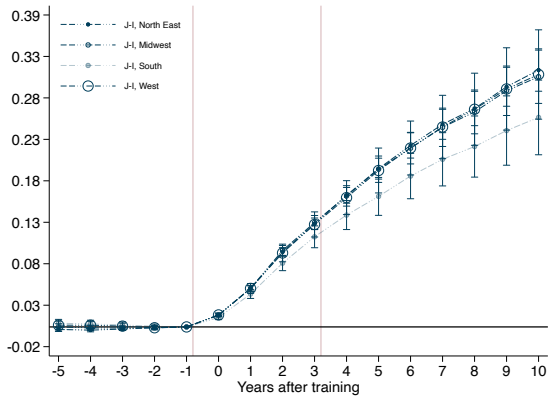
Panel F: ROA, share of managers leaving

Notes. Panels A to C show the treatment effects estimated separately by tertiles of the share of employees drafted during World War II. Panels D to F show the treatment effects estimated separately by tertiles of the share of top managers leaving the firm by 1955. To compute this variable, we compare the list of top executives in 1955 (or last year in which a firm is observed) to the list of top executives in period 0, the year of the TWI training. All regressions include the training variable in isolation, as well as fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The standard errors are clustered at the level of subdistricts and application windows.

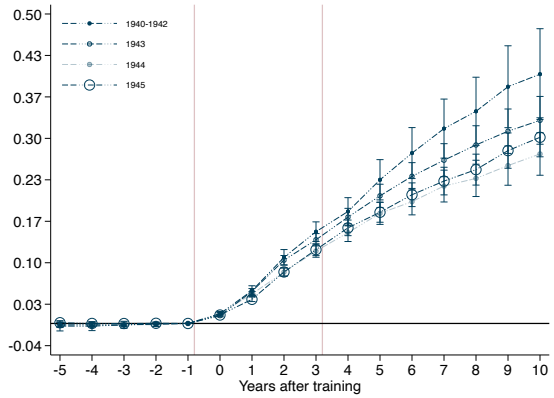
Figure A.10: Heterogeneous Effects by Sector, Region, Treatment Year, and Firm Size



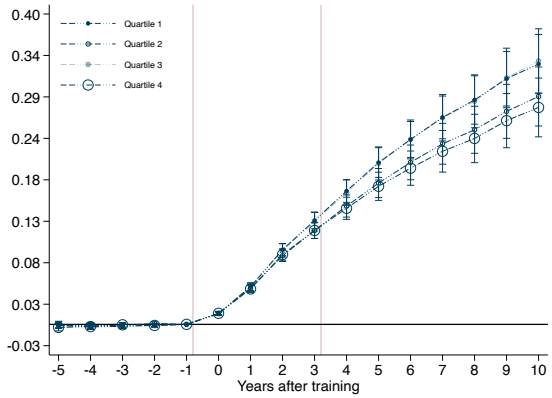
Panel A: Sectors



Panel B: Regions



Panel C: Treatment years



Panel D: Firm size

Notes. In all panels, the dependent variable is the log of TFPR, estimated using the Ackerberg, Caves, and Frazer (2015) method. Panel A shows the treatment effects estimated separately by sector. Panel B shows the treatment effects estimated separately by U.S. Census Bureau regions. Panel C shows the treatment effects estimated separately by training year. Panel D shows the treatment effects estimated separately by quartile of workforce size in period -1. All regressions also include fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The standard errors are clustered at the level of subdistricts and application windows.

Table A.1: List of 22 TWI Districts

District Name	States	Main Office Location
1) Upper New England	Maine; Massachusetts; Vermont; New Hampshire	Boston
2) Lower New England	Connecticut; Rhode Island	New Haven
3) Upstate New York	New York state (excluding Metropolitan New York)	New York
4) Metropolitan New York	Metropolitan New York	New York
5) New Jersey	New Jersey	Newark
6) Eastern Pennsylvania; Delaware	Eastern Pennsylvania; Delaware	Philadelphia
7) Maryland	Maryland	Baltimore
8) Atlantic Central	Virginia; North Carolina; South Carolina	Raleigh
9) South-Eastern States	Georgia; Florida; Alabama; Mississippi; Central and Eastern Tennessee	Atlanta
10) Ohio Valley	Southern Ohio; Southern West Virginia, Kentucky	Cincinnati
11) Western Pennsylvania	Western Pennsylvania (except Erie County); Norther West Virginia	Pittsburgh
12) Northern Ohio	Northern Ohio (expect Lucas County); Erie County (PA)	Cleveland
13) Michigan	Michigan; Lucas County (OH)	Detroit
14) Indiana	Indiana (except Lake and Porter Counties)	Indianapolis
15) Illinois	Illinois (except three counties adjacent to St. Louis, MO); South Wisconsin; Lake and Porter Counties (IN)	Chicago
16) North-Central States	North Wisconsin; Minnesota; North Dakota; South Dakota; Iowa; Nebraska	Minneapolis
17) South-Central States	Missouri; Kansas; Oklahoma; Arkansas; Western Tennessee; Madison, St. Clair, Monroe Counties (IL)	St. Louis
18) Gulf District	Texas; Louisiana	Houston
19) Mountain District	Colorado; Wyoming	Denver
20) Pacific Southwest	Southern California; Arizona; New Mexico	Los Angeles
21) Pacific Central	Northern California; Nevada; Utah	San Francisco
22) Pacific Northwest	Washington; Oregon; Idaho; Montana	Seattle

Notes. List of the 22 TWI districts with their headquarters location.

Table A.2: Balancing Tests for Applicant Firms By Treatment Year

	TWI	J-I	J-R	J-M	J-I and J-R	J-R and J-M	J-I and J-M	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: p-value of joint significance from regressions of training variables on firm characteristics in period -1								
Training in 1940-1942	0.36	0.97	0.41	0.60	0.41	0.77	0.40	0.85
Training in 1943	0.69	0.88	0.57	0.61	0.66	0.73	0.68	0.74
Training in 1944	0.08	0.56	0.46	0.39	0.95	0.35	0.81	0.48
Training in 1945	0.25	0.88	0.92	0.89	0.80	0.50	0.81	0.46
Panel B: p-value of joint significance from regressions of training variables on county characteristics in year 1940								
Training in 1940-1942	0.97	0.62	0.05	0.61	0.79	0.09	0.13	0.82
Training in 1943	0.02	0.51	0.38	0.07	0.55	0.67	0.17	0.12
Training in 1944	0.50	0.34	0.55	0.13	0.80	0.44	0.63	0.84
Training in 1945	0.15	0.95	0.84	0.52	0.84	0.22	0.14	0.18
Panel C: p-value of joint significance from regressions of training variables on county characteristics in year 1930								
Training in 1940-1942	0.82	0.70	0.04	0.79	0.87	0.27	0.05	0.67
Training in 1943	0.01	0.64	0.18	0.09	0.39	0.38	0.55	0.06
Training in 1944	0.68	0.32	0.42	0.06	0.45	0.73	0.86	0.84
Training in 1945	0.32	0.64	0.64	0.91	0.41	0.07	0.06	0.79
Panel D: p-value of joint significance from regressions of training variables on county characteristics in year 1920								
Training in 1940-1942	0.94	0.67	0.04	0.74	0.39	0.21	0.46	0.71
Training in 1943	0.14	0.50	0.21	0.02	0.49	0.39	0.39	0.14
Training in 1944	0.78	0.03	0.77	0.35	0.80	0.95	0.39	0.82
Training in 1945	0.19	0.70	0.60	0.70	0.36	0.34	0.14	0.27

Notes. Panel A shows the p-value of the test of joint significance of the coefficients of fifteen firm characteristics observed in period -1, distinguishing by training year. The variables are: the logs of sales, value added, number of employees, number of plants, foundation year, the value of inventory, capital, current assets, investments, number of workers' strikes, monetary compensation for workers' injuries, performance-based bonus payments, number of subsidiaries, as well as distance to the nearest railroad station and distance to the nearest port. The regressions also include fixed effects for county-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Panels B, C, and D show the p-value of the test of joint significance of the coefficients of several county characteristics measured in 1940, 1930, and 1920, distinguishing by training year. The county-level variables are imputed to firms based on their location. These regressions include: log population, log of the manufacturing value added, log number of manufacturing establishments, log number of manufacturing employees, log average manufacturing wage, log total expenses in manufacturing, log value of manufacturing production, farms per capita, unemployment share (available only in 1930 and 1940), population per square mile, share of male residents, share of black population, share of urban population, share of illiterate population (available only in 1920 and 1930). County data are from IPUMS NHGIS, www.nhgis.org. The regressions also include fixed effects for subdistrict-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Standard errors are clustered at the subdistrict-application window level.

Table A.3: Autocorrelation of the TWI Trainings

	TWI_t	J-I_t	J-R_t	J-M_t	J-I and J-R_t	J-I and J-M_t	J-R and J-M_t	All_t
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
TWI_{t-1}	0.028							
	(0.022)							
J-I_{t-1}		0.039						
		(0.028)						
J-R_{t-1}			-0.020**					
			(0.010)					
J-M_{t-1}				0.007				
				(0.028)				
J-I and J-R_{t-1}					0.002			
					(0.023)			
J-I and J-M_{t-1}						0.001		
						(0.022)		
J-R and J-M_{t-1}							0.007	
							(0.019)	
All_{t-1}								0.017
								(0.024)
Observations	1,873	1,873	1,873	1,873	1,873	1,873	1,873	1,873
R^2	0.001	0.002	0.001	0.001	0.001	0.001	0.001	0.001

Notes. This table shows the autocorrelation between the current and past share of firms that received different types of TWI trainings. The unit of observation is one of 364 subdistricts and one of 10 application windows. The standard errors are clustered at the subdistrict level. *** p<0.01, ** p<0.05, * p<0.1.

Table A.4: Changes in Firm Structure

	Log plants	Log employees	Log managers	Log subsidiaries	Log investment
	(1)	(2)	(3)	(4)	(5)
TWI x Period 1	0.0012***	0.0125***	0.0121	-0.0093	0.0620**
	(0.0003)	(0.0008)	(0.0080)	(0.0087)	(0.0263)
TWI x Period 5	0.0257***	0.0723***	0.0503***	0.0161*	0.1551***
	(0.0018)	(0.0016)	(0.0078)	(0.0089)	(0.0254)
TWI x Period 10	0.0622***	0.1268***	0.1084***	0.0496***	0.1678***
	(0.0030)	(0.0034)	(0.0083)	(0.0096)	(0.0278)
Observations	67,472	67,472	67,472	67,472	67,472
R^2	0.1443	0.1504	0.1416	0.1358	0.1493

Notes. This table shows the coefficients of the interactions between the training variable and three selected period dummies (out of 15 period dummies included). The omitted period is the year before the TWI training (period -1).The distance from the TWI intervention for the nontreated firms is imputed using the distance from the TWI intervention of the first participating firm in the same subdistrict and application window. The sample includes applicant firms that either received only one TWI training or no training at all. The standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

Table A.5: Effects of TWI Training on Responses to TWI Surveys

	J-I		J-R		J-M	
	Period 0	Period 3	Period 0	Period 3	Period 0	Period 3
	(1)	(2)	(3)	(4)	(5)	(6)
(1) Log machine repairs	-0.089***	-0.357***	0.053***	0.216***	0.053***	0.219***
	(0.001)	(0.006)	(0.001)	(0.004)	(0.001)	(0.004)
(2) Log workers' injuries	0.029***	0.118***	0.125***	0.029***	0.030***	0.121***
	(0.001)	(0.004)	(0.004)	(0.001)	(0.000)	(0.005)
(3) Register causes of breakdown	0.877***	0.877***	-0.019***	-0.018***	-0.020***	-0.016***
	(0.009)	(0.020)	(0.006)	(0.007)	(0.007)	(0.006)
(4) Job description for managers	0.001	0.001	0.876***	0.875***	-0.002	-0.002
	(0.001)	(0.002)	(0.006)	(0.019)	(0.002)	(0.003)
(5) Job description for workers	-0.019***	-0.020**	0.838***	0.826***	-0.015***	-0.014**
	(0.007)	(0.008)	(0.008)	(0.021)	(0.005)	(0.006)
(6) Training for workers	-0.025***	-0.030***	0.875***	0.875***	-0.015***	-0.013**
	(0.007)	(0.009)	(0.008)	(0.017)	(0.005)	(0.006)
(7) Bonus payment scheme	-0.008*	-0.009	0.897***	0.902***	-0.007*	-0.006
	(0.004)	(0.006)	(0.006)	(0.015)	(0.004)	(0.004)
(8) Suggestions from workers	0.001	-0.002	0.563***	0.552***	-0.002	-0.003
	(0.001)	(0.003)	(0.010)	(0.026)	(0.002)	(0.004)
(9) Log unused input	0.039***	0.157***	0.039***	0.158***	-0.089***	-0.348***
	(0.001)	(0.004)	(0.001)	(0.004)	(0.000)	(0.004)
(10) Production planning	-0.027***	-0.032***	-0.018**	-0.019**	0.823***	0.827***
	(0.008)	(0.009)	(0.007)	(0.008)	(0.009)	(0.023)
(11) Marketing	-0.008*	-0.007	-0.002	-0.001	0.840***	0.828***
	(0.004)	(0.006)	(0.002)	(0.003)	(0.008)	(0.020)
Observations	28,947	28,947	28,947	28,947	28,947	28,947

Notes. The data used in this table come from the surveys administered by the TWI administration only to firms that received at least one training. The responses were collected before the training, three months after the training, and then each year until 1945. Each row represents a separate regression whose dependent variable is one of the 11 management practices listed in the first column. The estimates are coefficients that measure the difference between each period and the period just before the implementation of the TWI program (period -1), separately for each TWI training. The regressions also include the coefficients for period 1 and 2, the treatment variables in isolation, as well as fixed effects for county-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Three dependent variables are continuous and transformed in logs: the number of machine repairs (1); the number of workers' injuries (2); and the value of unused input in the inventory (9). The other eight variables are dummies equal to 1 if the firm was implementing a given management practice: keeping records of machine breakdowns (3); providing clear list of tasks and expectations to managers (4) and workers (5); offering on-the-job training (6); implementing bonus payment schemes (7); allowing workers to provide suggestions (8); using formal production planning (10); having a formal marketing unit (11). Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

Table A.6: Complementarity Effects on Managerial Practices

	Log repairs	Log maintenance	Log injuries	Log bonus	Log strikes	Prob training	Log inventory	Log product lines	Prob marketing
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
J-I x Post	-0.038***	0.014***	-0.037***	-0.002	0.009***	-0.035***	0.009**	0.001	-0.026***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)	(0.004)	(0.004)	(0.003)
J-R x Post	-0.001	-0.001	-0.004	0.390***	-0.194***	0.543***	0.011***	0.004	-0.027***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.005)	(0.008)	(0.004)	(0.004)	(0.003)
J-M x Post	-0.004	-0.002	0.000	0.001	0.009***	-0.034***	-0.055***	0.334***	0.549***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.002)	(0.003)	(0.006)	(0.006)	(0.008)
J-I not alone x Post	0.006*	-0.013***	0.007	0.002	0.107***	0.015***	-0.011**	0.005	0.014***
	(0.003)	(0.002)	(0.005)	(0.004)	(0.003)	(0.004)	(0.005)	(0.005)	(0.004)
J-R not alone x Post	0.000	-0.007***	0.004	-0.012***	-0.101***	0.026***	-0.006	-0.006	0.011***
	(0.004)	(0.002)	(0.005)	(0.004)	(0.005)	(0.009)	(0.005)	(0.005)	(0.004)
J-M not alone x Post	0.004	0.006***	-0.004	-0.005*	0.107***	0.029***	-0.007	0.066***	0.027***
	(0.003)	(0.002)	(0.005)	(0.003)	(0.003)	(0.004)	(0.006)	(0.006)	(0.009)
Observations	134,288	134,288	134,288	134,288	134,288	134,288	134,288	134,288	134,288
R^2	0.101	0.098	0.101	0.210	0.177	0.615	0.395	0.552	0.627

Notes: This table shows the coefficients of the interactions between the training variables and a post-TWI dummy. All the dependent variables are logged with the exception of two dummy variables that measure whether firms reported expenditures for on-the-job training (column 6; mean=0.18) or for marketing activities (column 9; mean=0.17). J-x is one for all firms that received the J-x training, while J-x not alone is equal to one for firms that received the J-x module with another type of TWI training. The regressions also include the training variables in isolation, as well as fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. The sample includes all applicant firms. Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Selection of Firms Entering the Applicants' Supply Chain

	Log sales	TFPR	Log ROA	Log plants	Log employees	Log managers	Log subsidiaries
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
TWI	0.017***	0.313***	-0.018	0.099***	0.099***	0.111***	0.107***
	(0.001)	(0.022)	(0.033)	(0.012)	(0.009)	(0.027)	(0.011)
Observations	686	686	686	686	686	686	686
R-squared	0.533	0.407	0.120	0.266	0.337	0.148	0.287

Notes. This table shows cross-sectional correlations between the characteristics of upstream and downstream firms joining the supply chain of TWI applicants after the TWI program and the training variable. The only period considered is -1, the year before the entry in the supply chain of applicant firms. The sample includes 715 upstream and downstream firms that entered the supply chain of 660 TWI applicants only after the TWI program. The regressions also include fixed effects for district-sector combinations, the application window, and the number of days between the opening of the window and the firm application. Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Summary Statistics for 11,536 Nonapplicant War Contractors

	Nonapplicant firms				Difference with	Difference with
	Mean	St. Dev.	Min	Max	applicants	matched applicants
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Data from annual financial statements in 1939						
Plants	2.61	0.73	1	4	-3.47	-1.14
Employees	618.26	98.90	265	970	-420.06	10.19
Foundation year	1923	2.03	1916	1930	-8.01	-8.19
Agriculture	0.03	0.17	0	1	0.01	0.02
Manufacturing	0.86	0.35	0	1	-0.01	0.02
Transportation	0.09	0.29	0	1	0.01	-0.03
Services	0.02	0.15	0	1	0.01	-0.01
Sales	90.00	0.01	89.95	90.05	-101.78	2.02
Current assets	1.18	0.29	0.66	1.74	-17.22	-10.11
Total assets	3.64	0.35	2.95	4.39	-60.81	-34.45
TFPR	1.60	0.30	0.38	2.62	-0.15	-0.06
ROA	0.07	0.01	0.05	0.09	0.04	0.04
Inventory	0.54	0.07	0.28	0.84	-14.83	-8.56
Injuries	0.88	0.24	0.44	1.62	-9.52	-3.96
Repairs	1.35	0.26	0.90	1.80	-14.02	-5.74
Bonus payments	2.25	0.78	0.90	3.60	-6.02	-1.46
Panel B: Workforce data from replacement lists in 1941						
Share African-Americans	0.05	0.01	0.04	0.07	-0.01	-0.01
Share women	0.04	0.01	0.03	0.05	-0.01	-0.01
Years of education	7	1.05	3	11	-1.51	-1.52
Age of workforce	23.02	5.02	4	42	-5.96	-5.94
Panel C: WWII-related data from replacement lists and war contracts						
Share drafted employees (1942-1945)	0.20	0.09	0	0.35	-0.03	-0.03
Switched 3-digit SIC (1940-1945)	1	0	1	1	0	0
Switched 2-digit SIC (1940-1945)	0.54	0.50	0	1	-0.02	-0.01
Switched 1-digit SIC (1940-1945)	0.35	0.48	0	1	-0.02	-0.02
Number of contracts (1940-1945)	1.90	7.30	0	123	1.22	1.39
Value of contracts (1940-1945)	18.26	250.76	0	38,771	-6.85	-2.81

Notes. Summary statistics for 11,536 firms that applied to the TWI program and for which we found financial data (out of 13,818 total nonapplicants). Column 5 shows the average difference with respect to all applicants. Column 6 shows the average difference with respect to 101 applicants that were matched to nonapplicants. The matching was based on propensity scores, using a nearest-neighbor algorithm without replacement. The variables used for the matching were the number of plants and employees in 1939, sales and TFPR in 1939, distance from the nearest port and railroad station, and sector. A description of the variables is available under Table 1 and in Appendix xx.

Table A.9: Relationship with the Government

	Value contracts	Number contracts	Govt. TFPR	Non-govt. sales	Non-govt. TFPR	Postwar refunds
	(1)	(2)	(3)	(4)	(5)	(6)
TWI x Post	-0.1278	-0.0074	0.0284	0.0613***	0.0502***	
	(0.0906)	(0.0258)	(0.0301)	(0.0102)	(0.0095)	
TWI						-0.0000
						(0.0000)
Observations	7,207	7,207	6,861	16,636	16,636	20,388
R-squared	0.1882	0.2002	0.1808	0.1631	0.1453	0.1310

Notes. This table shows the effects of TWI training on variables that describe the relationship between applicant firms and the federal government. *Number contracts* and *Value contracts* are the log of the number and value of war contracts granted to each firm. *Govt. TFPR* is the TFPR that considers only sales to the government; *Non-govt. sales* is the log of sales to private firms; and *Non-govt. TFPR* is the TFPR that considers only private sales. These variables come from war contracts data and are therefore available only between 1940 and 1945. We can estimate only five periods (from -2 to 2) with a large number of firms. *Postwar Refunds* is the log of subsidies given by the government to war contractors to switch from military to civilian production after World War II. They are available only from 1946 to 1951, only after the end of the TWI program. Therefore, we need to estimate a single-difference specification, instead of the usual difference-in-differences event study. All regressions also include fixed effects for county-sector-period combinations, the application window, and the number of days between the opening of the window and the firm application. Standard errors are clustered at the level of subdistricts and application windows. *** p<0.01, ** p<0.05, * p<0.1.

C Data Collection and Variable Construction

Data Collection

The data collection targeted firms eligible to apply to the TWI program. In order to apply, firms had to be located in the United States and needed to win at least one war contract between June 1940 and August 1945. To identify the eligible companies, we used the tabulation of war supply contracts published by the Civilian Production Administration in 1946, which Dmitri Koustas digitized and kindly shared with us. This dataset includes information on all contracts for war supplies worth at least $50,000 and awarded between June, 1940 and September, 1945. Out of 25,627 companies with at least one contract in these five years, we excluded 234 firms located outside the United States, obtaining a list of 25,393 companies eligible to participate in the TWI program.[29] We then constructed a panel dataset gathering four different types of data.

TWI Monthly Program Development Reports (1940-1945). We retrieved and digitized the list of applicant firms from the monthly program development reports compiled by the TWI administration between August 22, 1940 and September 19, 1945. These reports are stored at the National Archives and Record Administration (Record Group 211, "Records of the War Manpower Commission [WMC]," 1936-1947, College Park, MD). Out of 25,393 eligible firms, we were able to find applications from 11,575 companies.[30] For each application, the monthly records indicate the application date, the district and subdistrict in which the applicant

[29] These companies, which were granted 772 contracts in total, were located in: Arabia (1), Argentina (22), Australia (4), Belgium (1), Bermuda (5), Brazil (5), British West Indies (6), Canada (123), Chile (3), China (2), Congo (1), Cuba (6), Denmark (1), Dominican Republic (2), Ecuador (1), Finland (1), India (5), Iran (1), Malaysia (1), Mexico, (9) Nicaragua (3), Panama (6), Peru (2), Philippines (3), Switzerland (8), Thailand (1), UK (7), and Venezuela (4).

[30] As noted in Brunet (2019), firms may appear in contract listings with numerous locations, which in some cases makes unclear whether there are multiple firms with the same name or one firm with multiple facilities.

firm was located, whether it eventually received the TWI training, in which of the J-modules it was trained, and the year in which each module was delivered. Moreover, we also retrieved and digitized the list of the instructors trained by the TWI service who delivered the J-modules to participant firms. For each instructor, we know their full name, whether they were employed by the government before the TWI program, whether they were hired full time or part-time, in which specific J-module they were trained, and to which subdistrict they were assigned. In total, the TWI instructors were 48,424.

Plant-level Surveys (1940-1945). In order to test the effectiveness of the program, the TWI administration conducted plant-level surveys in the firms that received the TWI training. Specifically, the surveys—stored at the National Archives and Record Administration (Record Group 211, "Records of the War Manpower Commission [WMC]," 1936-1947, College Park, MD)—indicate whether a plant was performing each of eleven managerial practices linked to the teachings of the TWI program. The plan managers were interviewed before the start of each J-module training, three months after the TWI training, and then each year thereafter until 1945. Overall, this dataset contain information about 38,241 surveyed establishments. We digitized these data and then we matched them to the list of applicant firms using name, municipality and state in which the firm's headquarters were located.

Replacement Schedules (1942-1945). After the Executive Order 9279 of December 5, 1942, firms in which at least one worker was drafted had to submit a replacement schedule to their regional Bureau of Manpower Utilization. In the schedules—stored at National Archives and Record Administration (Record Group 211, "Records of the War Manpower Commission [WMC]," 1936-1947, College Park, MD)—firms provided a description of their products, as well as a list of firms they were selling to and buying from. Moreover, they reported the composition of their labor force, specifically indicating the share of African-American workers and women, as well as the average years of education and age of all their employees. Finally, they had to list the names of drafted employees, their job titles, their relative ranking within the firm hierarchy, their age, their current Selective Service classification, their family status, their local board identity, and their draft order number, as well as the time needed to replace them with new workers (hence, then name "replacement schedules"). Through these replacement schedules, employers could also ask for exemptions from the draft for some categories of their workers. According to the Selective Training and Service Act of 1940, men between the ages of 18 and 45 were classified into four categories: (1) men available for training and service; (2) men deferred because of occupational status; (3) men deferred because of dependents; (4) men deferred by law or who were unfit for service. The Selective Service System, operating at a decentralized level through its 6,443 local boards, processed the exemption requests, mostly based on the information given by the draftees at the time of registering. Managers were usually deferred "in support of national health, safety, or interest" (category II-A). While the replacement schedules started being submitted in 1942, the first schedule contained data on both 1941 and 1942. Using firm name, municipality, and state, we searched for firms eligible for the TWI program in the replacement lists. Given the large size of war contractors, all of them had at least one worker drafted between 1940 and 1945. Therefore, we were able to match all applicant firms to the newly digitized dataset of replacement schedules.

Firm Performance (1935-1955). We retrieved information on the economic outcomes of U.S. war contractors from their balance sheets and income statements. To do so, we relied on the "Historical Annual Reports" collection of the Mergent Archives, which is composed of more than 1 million corporate financial statements collected directly from company archives, universities, libraries, and private collections.[31] In Summer 2016, we accessed this data thought the subscription of the University of California–Irvine. Specifically, we searched the UCI library's search engine for the name, municipality and state of each U.S. war contractor, collecting any document published between 1935 and 1955. For each firm and year found, we were able to download the annual financial reports a single pdf page at time after inserting a verification code for each downloaded page. If we did not find a firm before its first U.S. war contract, we used the foundation date to determine if this was due

We solve this ambiguity by checking how firms were listed in the TWI reports. Specifically, if we found two firms with the same name but in different locations listed as two separate entities receiving the TWI service, we consider them as two separate firms.

[31] https://www.ftserussell.com/data/mergent-archives.

to a missing statement or if the firm had not been founded yet. If we stopped finding a firm after its last U.S. war contract, we assumed that the firm had exited the market.. We were able to find at least one match for all the 11,575 applicants firms and for 11,536 out of 13,857 (83 percent) nonapplicants. We believe that the lower matching rate among nonapplicants is likely due to the smaller size. Even though there is not a formal threshold on firm size to be included in the Mergent Archives, in practice, publicly traded firms, firms issuing bonds, and larger firms are more likely to be included because it is relatively easier to find their balance sheets.

Variables Construction and Definition

In Table B.1, we provide a list of all the variables used in the paper with their definitions and data sources. When needed, we also provide additional details on the variable construction. All the monetary values have been reevaluated to 2020 USD.

Definition of Sector and Identification of Firms in the Supply Chain

The annual financial reports do not include information on a firm's sector. To retrieve this piece of information, we rely on the replacement schedules. In fact, firms had to report the products not covered by a war contract for the year 1941, the year before the draft started. We therefore classify these products using the Standard Industrial Classification (SIC) created for the first time in U.S. history in 1939 for the manufacturing products and in 1940 for the non-manufacturing products at 1-, 2-, and 3-digit levels.[32] Specifically, we find the classification of manufacturing products in Volume I, part 3—"Alphabetic Index of Products, Establishments and Processes, 1939," and the classification of the products in Volume II, Part 3—"Alphabetic Index of Products, Establishments, and Services, 1940," published by the U.S. Office of Management and Budget in 1941 under the title of "Standard Industrial Classification Manual."

We then use the same classifications to obtain the 1-, 2-, and 3-digit SIC codes of each firm's war products. We use this information to construct three variables: *Switched 3-digit SIC*, an indicator that equals to 1 for firms in which at least one war product had a different 3-digit SIC code than the peacetime products listed in 1941; *Switched 2-digit SIC*, an indicator that equals to 1 for firms in which at least one war product had a different 2-digit SIC code than the peacetime products listed in 1941; *Switched 1-digit SIC*, an indicator that equals to 1 for firms in which at least one war product had a different 1-digit SIC code than the peacetime products listed in 1941.

In the replacement schedules, firms also had to indicate the name of the companies with which they had an existing contract, as well as the list of products bought from and sold to them. We use this information to construct a list of upstream and downstream firms for the U.S. war contractors. First, we use the SIC codes to classify the 2-digit industry of the products sold to and bought from these companies. Then, we use the 1947 input-output tables to assess whether these products were upstream and downstream, relative to the products of the U.S. war contractor.[33] To avoid any possible ambiguity, we keep only firms for which all the products had the same 2-digit industry. Out of 3,465 companies listed in the replacement schedules, we exclude 341 firms. Out of the remaining 3,124 firms, we were able to find financial statements for 1,816 of them, which represent the sample we use for the analysis in Section 6.1.

[32] While the SIC categorizes products up to 4 digits, the product description is not accurate enough for such a detailed classification.

[33] The 1947 input-output tables have been digitized by Soltan (2019) and are available at the following link: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2FMMHRU7 (downloaded in September 2018).

Table B.1: List, Definition, and Sources of Variables

Variable	Definition	Level, Source and Years of Coverage	Notes
Plants	Number of firm plants	Firm-year, Financial Statement, 1935-1955	
Employees	Number of employees	Firm-year, Financial Statement, 1935-1955	
Foundation year	Year of firm's foundation	Firm-year, Financial Statement, 1935-1955	
Sector	Sector in which firm operated	Firm, Replacement Schedules, imputed in 1941	See Section C.
Sales	Annual revenues from sales	Firm-year, Financial Statement, 1935-1955	
Current assets	Assets that can be converted to cash within one year or less	Firm-year, Financial Statement, 1935-1955	
Total assets	Total amount of assets	Firm-year, Financial Statement, 1935-1955	
TFPR	Total factor productivity revenue	Firm-year, Financial Statement, calculated in 1935-1955	Authors' calculation following Ackerberg, Caves, and Frazer (2015)'s methodology (see Figure A.3)
ROA	Ratio between profits and fixed gross assets	Firm-year, Financial Statement, calculated in 1935-1955	Authors' calculation
Fixed gross assets	Fixed gross asset is the value of land, buildings, and machines owned by the firm. To estimate its replacement costs from the historical cost reported in the financial statement, we used the R^G factor suggested by Balakrishnan et al. (2000).	Firm-year, Financial Statement, calculated in 1935-1955	$R^G = \frac{[(1+g)^{\tau+1}-1](1+\pi)^{\tau}[(1+g)(1+\pi)-1]}{g\{[(1+g)(1+\pi)]^{\tau+1}-1\}}$ where τ is the average life of machines (assumed to be 15 years), π is the average capital price $\frac{P_t}{P_{t-1}}$ from 1935 to 1955 (equal to 1.00476667), and g is the (assumed constant) real GDP growth rate from 1935 to 1955 (equal to 1.0589524).
Span of control	Ratio of other employees and managers	Firm-year, Financial Statement, 1935-1955	Authors' calculation
Inventory	Value of inventory at cost	Firm-year, Financial Statement, 1935-1955	
Maintenance	Cost of machine maintenance	Firm-year, Financial Statement, 1935-1955	
Injuries	Cost of workers' injuries	Firm-year, Financial Statement, 1935-1955	
Repairs	Cost of machine repairs	Firm-year, Financial Statement, 1935-1955	
Bonus payments	Cost of bonus paid to workers	Firm-year, Financial Statement, 1935-1955	
Product lines	Number of product lines	Firm-year, Financial Statement, 1935-1955	
Prob training	=1 if firm reported spending for on-the-job training	Firm-year, Financial Statement, 1935-1955	
Prob marketing	=1 if firm reported spending for marketing	Firm-year, Financial Statement, 1935-1955	
Strikes	Number of strikes	Firm-year, Work Stoppages Caused by Labor Management Disputes, BLS, 1935-1955	
Share African-Americans	% of African-American workers	Firm-year, Replacement Schedules, 1941-1945	
Share women	% of women in workforce	Firm-year, Replacement Schedules, 1941-1945	
Years of education	Average years of education of workforce	Firm-year, Replacement Schedules, 1941-1945	
Age of workforce	Average age of workforce	Firm-year, Replacement Schedules, 1941-1945	
Share drafted employees (1942-1945)	% of drafted employees	Firm, Replacement Schedules, 1942-1945	
Switched 3-digit SIC	Indicator if at least one war product was different from products in 1941 in 3-digit 1939 SIC codes	Firm, Civilian Production Administration and Replacement Schedules, imputed for the period 1940-1945	See Section C.
Switched 2-digit SIC	Indicator if at least one war product was different from products in 1941 in 2-digit 1939 SIC codes	Firm, Civilian Production Administration and Replacement Schedules, imputed for the period 1940-1945	See Section C.
Switched 1-digit SIC	Indicator if at least one war product was different from products in 1941 in 1-digit 1939 SIC codes	Firm, Civilian Production Administration and Replacement Schedules, imputed for the period 1940-1945	See Section C.
Number of contracts	Number of war contracts	Firm-year, Civilian Production Administration, 1940-1945	Aggregated at the firm-year level from the original contract-level data
Value of contracts	Value of contracts	Firm-year, Civilian Production Administration, 1940-1945	Aggregated at the firm-year level from the original contract-level data
Govt. TFPR	TFPR calculated with share of input and output used for contracts with government	Firm-year, Civilian Production Administration, 1940-1945	Share of input and output used for contracts with government is value contracts divided by annual sales
Non-govt. TFPR	TFPR calculated with share of input and output not used for contracts with government	Firm-year, Civilian Production Administration, 1940-1945	Share of input and output used for contracts with government is value contracts divided by annual sales
Post-WWII refunds	Subsidies given to firms to go back to civil production	Firm-year, Financial Statement, 1946-1951	

D Results on TWI Instructors' Assignments

In this section of the appendix, we show some results that point to the fact that the assignment of TWI instructors to subdistricts and application windows influenced the type of training that firms received. Moreover, this assignment was not driven by the applicants' characteristics.

The TWI administration would have certainly preferred to send enough instructors to train all war contractors in all three J-modules, but several bottlenecks made this goal unreachable. First, the TWI staff was not large enough to prepare enough TWI instructors for in-plant training. At its peak in July 1944, the TWI administration had 415 staff members to serve 100,000 plants employing 19 million workers (Dooley, 1945, p. 14). During most of WWII, the size of the TWI staff was much smaller and the majority of staff members was working only part-time for the program. Second, the lack of manpower was compounded by the fact that both the TWI Institutes (the courses to certify instructors) and the TWI training itself had to be carried out in small classes of at most ten students, severely limiting scalability. Third, as discussed in Section 2.3, TWI instructors were specialists in the sense that they could teach only one J-module. This decision sped up the delivery of the training, but meant that some applicant firms could not receive one or more J-modules because there were not enough instructors that could teach them in their subdistrict and application window. To conclude, a small staff, as well as instructors with inflexible teaching skills, created variation across subdistricts and application windows in the number of firms eventually trained and the type of trainings they received.

The tables in this section shows the following results. First, within a subdistrict and application window, the share of instructors who could teach a given J-module is positively and strongly correlated with the probability that firms were trained in the same J-module (Table C.1). Second, the share of instructors who could teach a given J-module is not correlated with the baseline characteristics of firms that applied in the same subdistrict and application window (Table C.2, Panel A). Similarly, there is no correlation between the share of instructors and the characteristics of counties in 1940, 1930, or 1920 (Table C.2, , Panels B-D). Fourth, within a given subdistrict, there is no autocorrelation between the share of instructors who could teach a given J-module between two adjacent application windows (Table C.3).

In short, the assignment of TWI instructors to subdistricts and application windows did not depend on the characteristics of applicants firms, the characteristics of the counties, nor prior assignments. These results are consistent with the idea that the TWI administration did not act strategically in assigning their resources to applicant firms.

Based on these findings, we can modify the baseline equation (1). Specifically, we can instrument the training variables with the share of instructors who could teach different J-modules assigned to the subdistrict and application window in which firm i applied to the program. Table C.1 showed that the instruments are not weak. Tables C.2 and C.3 suggested that the instruments are unlikely to violate the exclusion restriction. As expected, the IV and OLS estimates are close (Table C.4).

Table C.1: Correlation Between TWI Training and Instructors' Composition

	J-I first	J-R first	J-M first	J-I + J-R	J-R + J-I	J-R + J-M	J-M + J-R	J-I + J-M	J-M + J-I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Max J-I	0.215***								
	(0.009)								
Max J-R		0.197***							
		(0.010)							
Max J-M			0.192***						
			(0.008)						
Max J-I + J-R				0.136***					
				(0.013)					
Max J-R + J-I					0.128***				
					(0.011)				
Max J-R + J-M						0.166***			
						(0.016)			
Max J-M + J-R							0.080***		
							(0.013)		
Max J-I + J-M								0.152***	
								(0.014)	
Max J-M + J-I									0.179***
									(0.009)
Observations	10,735	10,735	10,735	10,735	10,735	10,735	10,735	10,735	10,735
R^2	0.189	0.192	0.188	0.146	0.147	0.146	0.121	0.135	0.150

Notes. This table shows the correlations between the firm-level TWI training and the share of TWI instructors assigned to a firm's subdistrict and application window. The dependent variables are the following dummies: J-I first = 1 for firms that first received the Job-Instructions training; J-R first = 1 for firms that first received the Job-Relations training; J-M first = 1 for firms that first received the Job-Methods training; J-I + J-R = 1 for firms that received first the J-I training and then the J-R training; J-R + J-I = 1 for firms that received first the J-R training and then the J-I training; J-R + J-M = 1 for firms that received first the J-R training and then the J-M training; J-M + J-R = 1 for firms that received first the J-M training and then the J-R training; J-I + J-M = 1 for firms that received first the J-I training and then the J-M training; J-M + J-I = 1 for firms that received first the J-M training and then the J-I training. The independent variables measure the following: Max J-x = 1 for firms that applied in subdistricts and application windows in which the share of instructors for training J-x (I, R, or M) was the highest; Max J-x + J-y = 1 for firms that applied in subdistricts and application windows in which the share of instructors for training J-x (I, R, or M) was the highest and the share of instructors for training J-y ($y \neq x$) was the second highest. The regressions also include fixed effects for county-sector pairs, the application window, and the number of days between the opening of the window and the firm application. The standard errors are clustered at the subdistrict-application window level. *** p<0.01, ** p<0.05, * p<0.1.

Table C.2: Correlation Between Firm and County Variables and TWI Instructors

	Max J-I	Max J-R	Max J-M	Max J-I + J-R	Max J-R + J-I	Max J-R + J-M	Max J-M + J-R	Max J-I + J-M	Max J-M + J-I
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Regressions of training variables on firm characteristics in period -1									
p-value of joint significance	0.98	0.98	0.37	0.76	0.90	0.96	0.51	0.97	0.64
Observations	10,735	10,735	10,735	10,735	10,735	10,735	10,735	10,735	10,735
Panel B: Regressions of training variables on county characteristics in year 1940									
p-value of joint significance	0.01	0.14	0.01	0.01	0.71	0.07	0.86	0.23	0.12
Observations	10,745	10,745	10,745	10,745	10,745	10,745	10,745	10,745	10,745
Panel C: Regressions of training variables on county characteristics in year 1930									
p-value of joint significance	0.05	0.19	0.14	0.05	0.83	0.36	0.66	0.07	0.71
Observations	10,859	10,859	10,859	10,859	10,859	10,859	10,859	10,859	10,859
Panel D: Regressions of training variables on county characteristics in year 1920									
p-value of joint significance	0.04	0.18	0.12	0.01	0.57	0.09	0.72	0.40	0.85
Observations	10,873	10,873	10,873	10,873	10,873	10,873	10,873	10,873	10,873

Notes. Panel A shows the *p*-value of the test of joint significance of the coefficients of fifteen firm characteristics observed in period -1. The variables are: the logs of sales, value added, number of employees, number of plants, foundation year, the value of inventory, capital, current assets, investments, number of workers' strikes, monetary compensation for workers' injuries, performance-based bonus payments, number of subsidiaries, as well as distance to the nearest railroad station and distance to the nearest port. The regressions also include fixed effects for county-sector pairs, the application window, and the number of days between the opening of the window and the firm application. Panels B, C, and D show the *p*-value of the test of joint significance of the coefficients of several county characteristics measured in 1940, 1930, and 1920, respectively. The county-level variables are imputed to firms based on their location. These regressions include: log population, log of the manufacturing value added, log number of manufacturing establishments, log number of manufacturing employees, log average manufacturing wage, log total expenses in manufacturing, log value of manufacturing production, farms per capita, unemployment share (available only in 1930 and 1940), population per square mile, share of male residents, share of black population, share of urban population, share of illiterate population (available only in 1920 and 1930). County data are from IPUMS NHGIS, www.nhgis.org. The regressions also include fixed effects for subdistrict-sector pairs, the application window, and the number of days between the opening of the window and the firm application. The description of the dependent variables is in the notes of Table C.1. Standard errors are clustered at the subdistrict-application window level.

Table C.3: Autocorrelation between Current and Past TWI Instructors

	Max J-I$_t$	Max J-R$_t$	Max J-M$_t$	Max J-I + J-R$_t$	Max J-R + J-I$_t$	Max J-R + J-M$_t$	Max J-M + J-R$_t$	Max J-I + J-M$_t$	Max J-M + J-I$_t$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Max J-I$_{t-1}$	0.002								
	(0.022)								
Max J-R$_{t-1}$		0.019							
		(0.025)							
Max J-M$_{t-1}$			-0.002						
			(0.022)						
Max J-I + J-R$_{t-1}$				-0.020					
				(0.023)					
Max J-R + J-I$_{t-1}$					0.016				
					(0.026)				
Max J-R + J-M$_{t-1}$						0.046*			
						(0.026)			
Max J-M + J-R$_{t-1}$							-0.037*		
							(0.020)		
Max J-I + J-M$_{t-1}$								0.012	
								(0.023)	
Max J-M + J-I$_{t-1}$									-0.006
									(0.023)
Observations	1,873	1,873	1,873	1,873	1,873	1,873	1,873	1,873	1,873
R^2	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001	0.002	0.001	< 0.001	< 0.001

Notes. The table shows the autocorrelation between the current and past share of instructors for different types of TWI trainings. The unit of observation is one of 364 subdistricts and one of 10 application windows. The description of the dependent variables is in the notes of Table C.1. The standard errors are clustered at the subdistrict level. *** p<0.01, ** p<0.05, * p<0.1.

Table C.4: IV and OLS Event Studies

	TFPR		Sales		ROA	
	(1)	(2)	(3)	(4)	(5)	(6)
J-I x Post TWI	0.0810***	0.0706	0.1197***	0.1854**	-0.0053	-0.0252
	(0.0040)	(0.0621)	(0.0073)	(0.0857)	(0.0044)	(0.0558)
J-R x Post TWI	0.3187***	0.3142***	0.1545***	0.2625***	0.1469***	0.1352**
	(0.0062)	(0.0606)	(0.0083)	(0.0831)	(0.0046)	(0.0544)
J-M x Post TWI	0.1372***	0.1093***	0.1126***	0.1781***	0.0517***	0.0236
	(0.0043)	(0.0396)	(0.0069)	(0.0542)	(0.0055)	(0.0355)
Specification	OLS	IV	OLS	IV	OLS	IV
Observations	67,472	67,472	67,472	67,472	67,472	67,472

Notes. This table shows the coefficients of the interactions between the training variables and a post-TWI dummy variable. In the IV specifications, we instrument for the TWI training variables using three dummy variables that measure the allocation of TWI instructors to each subdistrict and application window: Max J-x = 1 for firms that applied in subdistricts and application windows in which the share of instructors for training J-x (I, R, or M) was the highest. The sample includes applicant firms that either received only one TWI training or no training at all. The standard errors are clustered at the subdistrict-application window level. *** p<0.01, ** p<0.05, * p<0.1.