Can Sourcing Help Enforce Global Labor Standards?
Evidence from the Gap Inc Supply Chain

Matthew Amengual
Greg Distelhorst

October 4, 2019

Multinational firms with global supply chains manage the labor and environmental practices of their supplier firms. Does this "private regulation" cause social compliance to improve? How does the organization of these programs shape their efficacy? This study uses a regression discontinuity design to estimate the causal effects of private regulation on labor compliance, analyzing over one thousand factories supplying the multinational retailer Gap Inc. This retailer's 2016 change in management practices allows us to study private regulation when first decoupled from and later aligned with the activity of the sourcing department, which controls allocation of orders to factories. Before alignment we estimate precise, near-zero effects of failing audits on future social compliance. However, after the alignment of sourcing with supplier responsibility, failing an audit caused factory compliance to improve by 0.8 standard deviations and reduced the probability of future failure by 22 percentage points. Independent labor compliance data from the ILO/IFC Better Work program shows similarly large effects. Longer-term suppliers improved most, consistent with the hypothesis that committed commercial relationships enhance the credibility of demands to improve labor standards. Private regulation did cause improved labor compliance. Yet it only did so after aligning sourcing activity with the management of social responsibility.

Keywords: private regulation, decoupling, supply chain, social responsibility, social performance, labor standards, compliance.

Global supply chains pose risks to the reputation and financial performance of multinational enterprises. Recent media reports link factories and farms in global supply chains to environmental damage, industrial accidents, wage theft, and even modern slavery.[1] These abuses open the door to activist campaigns that target multinational enterprises, threaten their brands and reputations, and impose financial costs (Baron 2001; King and Soule 2007; Bartley and Child 2011).

To manage these risks, many multinationals manage not only the operational performance of their suppliers but also their *social* performance. Over 90% of Fortune's top 250 global corporations have adopted supplier codes of conduct,[2] and a majority of all publicly listed corporations in the food, textile, and wood products industries manage social responsibility in their supply chains (Thorlakson, de Zegher and Lambin 2018). This "private regulation" of global labor and environmental practices has attracted scholarly interest in strategic management (Vogel 2010; Short, Toffel and Hugill 2016), operations (Lee and Tang 2017; Jira and Toffel 2013; Kalkanci, Ang and Plambeck 2016; Liu et al. 2019; Kalkanci and Plambeck 2019), and organizations (Bird, Short and Toffel 2019), as well as economics (Harrison and Scorse 2010; Tanaka 2017; Boudreau 2018), sociology (Bartley 2007; Seidman 2007), and political science (Anner 2012; Locke 2013).

This study addresses two questions about private regulation in supply chains. First, does private regulatory activity cause compliance to improve? This question is at the center of the longstanding tension between private regulation optimists, who note its potential to raise standards in emerging markets (Fung, O'Rourke and Sabel 2001; Elliott and Freeman 2003; Lee and Tang 2017), and more skeptical studies pointing to violations of basic standards even in supply chains subject to private regulation (Barrientos and Smith 2007; Locke 2013; Vogel 2010; Anner 2012). Despite nearly two decades of research, the causal effects of private regulation have remained elusive.

Second, is private regulation more effective when companies align the activity of their sourcing departments—which manage commercial relationships with suppliers— with regulatory goals? Private regulation is often led by specialized social responsibility managers. Their teams monitor suppliers and push them to comply with labor and environmental standards, but they do not have direct control over purchasing decisions. Scholars and practitioners alike argue that breaking down barriers between sourcing and social responsibility could make private regulation more effective (Oxfam 2010; Anner, Bair

---

[1] Amazon Deforestation, Once Tamed, Comes Roaring Back. *The New York Times.* https://tinyurl.com/zqs8atw. More Than 300 Killed in Pakistani Factory Fires. *The New York Times.* https://tinyurl.com/ybsnf2ko. Building Collapse in Bangladesh Leaves Scores Dead. *The New York Times.* https://tinyurl.com/yaln4pgo. Human Rights Watch. "Thailand: Forced Labor, Trafficking Persist in Fishing Fleets," January 23, 2018. Worker Rights Consortium "Stealing from the Poor: Wage Theft in the Haitian Apparel Industry" 2013.

[2] KPMG International. 2008. "KPMG International Survey of Corporate Responsibility Reporting 2008."

2

and Blasi 2013; ILO 2017), but no research to date has been able to compare the impacts of private regulation when decoupled from and aligned with sourcing.

This study addresses these questions by analyzing nine years of compliance audits, sourcing records, and administrative data from over one thousand factories in the supply chain of Gap Inc, a large multinational apparel retailer. To estimate the causal effect of supplier responsibility activities, we exploit a sharp discontinuity in the assignment of labor compliance ratings. Failing a labor audit triggered different private regulatory actions in different periods of our study. In the earlier period of our study, a failing rating communicated buyer displeasure and prompted social responsibility managers to allocate more support for remediation to the failing factory. By contrast, since 2016 failing has additionally triggered communication from the sourcing department, raising the possibility that the factory could be terminated from the supply chain if it did not improve. This change allows us to study the effects of private regulation before and after aligning the behavior of the sourcing department with the goals of the supplier responsibility program.

Before alignment with sourcing, we find that failing a compliance audit had a precisely estimated null effect on future compliance. Yet after breaking down the silos separating social responsibility and sourcing managers, failing caused factory compliance scores to improve by 0.8 [0.2, 1.4] standard deviations, increasing the probability of passing the next audit by 22 percentage points. To address the possibility of bias in the buyer's in-house compliance audits, we validate these findings using independent audits conducted by Better Work, a program of the International Labor Organization and the World Bank International Finance Corporation. Finally, we analyze heterogenous treatment effects by suppliers with longer versus shorter-term commercial relationships with the buyer, finding that the combination of aligned sourcing and longer-term commercial relationship yields the greatest improvement.

These findings have implications for scholarly debates, management practice, and public policy. Our results address a longstanding debate on the efficacy of private regulation by illustrating the limits of private regulation as it is widely practiced, and its potential for greater impact when better aligned with other supply chain management activities. Our analysis of the consequences of policies adopted in response to stakeholder pressure contributes to scholarship on both private politics (Baron 2003) and means–ends decoupling in organizations (Bromley and Powell 2012). Managers can use these findings to design supplier responsibility programs—widely adopted in several industries—for greater impact. Finally, this study suggests that policymakers can distinguish more and less effective forms of private regulation by investigating the links between sourcing and supplier responsibility programs. This question is increasingly relevant in public policy as new laws such as the UK Modern Slavery Act and the French Law on Duty of Care require multinational corporations to take responsibility for social practices in their supply chains.

3

Private Regulation of Supply Chains and its Impact

The private regulation of global supply chains emerged from private politics, in which "interest and activist groups attempt to influence economic activity directly without reliance on public institutions" (Baron 2001, 7). Social movements concerned with worker rights, deforestation, conflict minerals, marine mammals, and other issues have targeted large firms in advanced economies with the goal of influencing behavior in hundreds or even thousands of their supplier firms, often in emerging markets. These "proxy targeting" campaigns (Walker, Martin and McCarthy 2008) pressed multinational enterprises to bring their suppliers' practices in line with international standards of socially responsible production. They prompted many multinational firms to develop programs to regulate the social and environmental performance of their supplier factories, forests, or farms (Bartley 2007). Private regulation has taken many organizational forms, including corporate codes of supplier conduct, multistakeholder initiatives, and social and environmental certification schemes (Bartley 2003; Cashore, Auld and Newsom 2004).

Private regulation in supply chains is premised on the idea that buying firms can change the labor and environmental practices of their suppliers. The common suite of private regulatory activity includes monitoring suppliers via compliance audits, documenting violations, assisting suppliers in identifying the root causes of violations, and developing "corrective action" or "remediation" plans, which detail how the root causes of each violation will be addressed. Because developing country firms exhibit different management practices from advanced economy firms (Bloom and Van Reenen 2010), private regulatory engagement may transfer knowledge and practices that enable compliance (Locke, Amengual and Mangla 2009). These activities—identifying problems, providing advice, and monitoring progress—may be more effective when buyers and suppliers can better cooperate with one another (Hugill, Short and Toffel 2016), especially in the context of a long-term business relationship (Locke 2013).

Research on private regulation in supply chains has produced important stylized facts, but it has struggled to estimate private regulation's impact on regulated firms. One consistent finding is that supplier firms exposed to the common suite of private regulatory activity remain far from fully compliant (Frenkel 2001; Barrientos and Smith 2007; Locke, Qin and Brause 2007; Anner 2012; Toffel, Short and Ouellet 2015; Bird, Short and Toffel 2019). Although most research focuses on the apparel and footwear industries, similar patterns emerge in both electronics (Raj-Reichert 2013; Distelhorst et al. 2015; Yang and Gallagher 2017) and agricultural products (Riisgaard 2009; Coslovsky and Locke 2013; Dietz, Grabs and Estrella Chong 2019).

Scholars offer several explanations for why private regulation fails to generate full compliance. First, intense competition in globalized industries may constrain compliance with international labor and environmental standards. Buyers and suppliers face pressure

to keep costs low and meet volatile consumer demand. In industries such as apparel and electronics, buyers capture a large share of value and suppliers face narrow margins (Gereffi 1999; Locke and Samel 2018; Anner Forthcoming). Suppliers under such pressures may adopt practices to lower costs that undermine labor compliance (Vogel 2005). For example, suppliers use piece-rate pay systems to support higher labor productivity, but these systems can render private regulatory systems less effective (Bird, Short and Toffel 2019). Furthermore, volatile customer demand can lead to production spikes, managed in part through excessive overtime hours, followed by low demand periods in which suppliers must shed employees (Locke and Samel 2018; Anner Forthcoming). Recognition of these pressures has led civil society groups to conclude that current industry structures and business models undermine private regulation (Reinecke et al. 2019).

Second, suppliers often operate in countries with weak regulatory enforcement and poor freedom of association for workers. In the absence of effective government regulation or well-organized trade unions, suppliers can cut costs in ways that violate laws or international social and environmental standards with few negative consequences. Private regulation is strengthened when government regulation is also strong, but many developing country governments lack the capacity or will to enforce their own regulations (Amengual 2010; Locke, Rissing and Pal 2013). Empirical research shows that supplier compliance is lower in countries with weaker laws and fewer local institutional pressures to follow standards (Toffel, Short and Ouellet 2015; Distelhorst et al. 2015), supporting the argument that global standards simply cannot transcend local constraints on compliance (Bartley 2018).

However, these pressures and the well-documented failure to converge at full compliance do not mean that private regulation has no effect. Indeed, longitudinal studies of compliance outcomes show that suppliers do become more compliant over repeated audits (Barrientos and Smith 2007; Coslovsky and Locke 2013; Locke 2013; Bird, Short and Toffel 2019). This evidence is consistent with the claim that private regulation increases compliance, but the same trends could also arise from a range of confounding factors other than private regulatory activity.

Evidence on the causal effects of private regulatory activity would advance this debate about a widely adopted management practice. Yet only a handful of previous studies measure supplier improvement against plausible counterfactuals. Harrison and Scorse (2010) used a combination of sectoral and geographic variation in exposure to private regulatory pressures to show that activist campaigns increased wages in Indonesia. Boudreau (2018) engineered experimental manipulation of a firm-led intervention to promote worker safety committees in Bangladeshi exporters. Finally, Thorlakson, Hainmueller and Lambin (2018) combined matching and difference-in-differences approaches to approximate counterfactual outcomes in South African farms supplying a major grocery chain. Despite these advances, there remains little evidence on the causal effects of widely

adopted private regulatory activities: monitoring compliance with codes of conduct, rating compliance, and supporting improvement. We therefore lack an understanding of which combination of these activities, if any, makes private regulation effective.

Private Regulation Under Decoupled and Aligned Sourcing

In addition to the competitive pressures and institutional explanations, the *organization* of private regulatory activity offers another possible explanation for its limited efficacy. As described above, multinationals adopted supply chain responsibility programs in response to stakeholder pressure. When firms adopt formal structures to maintain legitimacy in the eyes of stakeholders, they may decouple these structures from core organizational practices (Meyer and Rowan 1977). Decoupling includes the symbolic adoption of legitimacy-enhancing policies with no implementation in practice (i.e. policy–practice decoupling). However, even when policies are faithfully implemented we may still observe means–ends decoupling: implementation that does not produce the ostensibly desired outcome. Means–ends decoupling can occur when the implementation of legitimacy-enhancing policies is relegated to "silos within organizations" (Bromley and Powell 2012, 499), separate from other parts of the organization that are essential for achieving the policy's stated goal.

The sourcing department is one part of the organization that appears important to the success of private regulation in global supply chains. Misalignment of sourcing activity and private regulatory activity might influence suppliers' expectations about the consequences for failing to comply. Sourcing managers decide which suppliers to contract with, how much volume to purchase, and when to terminate commercial relationships. These activities are important to managing the performance of suppliers, yet research suggests that the sourcing department is often "internally buffered" (Meyer and Rowan 1977) from social responsibility programs (Locke 2013). Even sophisticated supply chain responsibility programs may be poorly aligned with sourcing activity (Amengual, Distelhorst and Tobin 2019). Surveys of suppliers show that demands from sourcing departments often conflict with demands from buyers' social responsibility staff (ILO 2017; Dickson 2018), and surveys of buyers show that fewer than 30% terminate contracts with suppliers that repeatedly violate standards (Porteous, Rammohan and Lee 2015).

Aligning sourcing with supplier responsibility could make private regulation more effective if suppliers are motivated to comply by the value of the future business relationship.[3] Sourcing managers, rather than social responsibility managers, decide which business

---

[3]This argument builds on theories of relational contracting (Gibbons 2005). A relational contracting approach is appropriate because supply chain responsibility is managed through informal rather than formal contracts and third-party dispute settlement. In this sense, private regulation mirrors what scholars have observed in the management of quality, price, and delivery in supply chains (Sako 2004; Helper and Henderson 2014; Macchiavello and Morjaria 2015).

relationships with suppliers will continue and which will end. By aligning the activities of the sourcing department with supplier responsibility, suppliers' compliance with labor and environmental standards can be directly linked to the future of the business relationship. In the absence of congruent action by different units within the buying firm, suppliers may infer that the business relationship does not depend on their social performance and ignore requests from social responsibility managers to improve their practices.

Yet there is no guarantee that using the sourcing department to link the future of the business relationship to supplier labor and environmental standards would overcome the barriers to effective private regulation identified in previous research. Aligning sourcing with supplier responsibility does nothing to address the institutional factors that limit private regulation, including ineffective government regulators and anemic civil society (Toffel, Short and Ouellet 2015; Distelhorst et al. 2015; Bartley 2018). Nor would it reduce the pressures that suppliers face in highly competitive industries, which may bind their labor and environmental practices (Gereffi 1999; Vogel 2005; Anner Forthcoming).

Long-term Commercial Relationships and Private Regulation

In addition to the organizational dimensions discussed above, we also study how the length of the buyer–supplier commercial relationship conditions the efficacy of private regulation. Scholars have repeatedly pointed to the importance of embedding private regulation within long-term commercial relationships. Locke (2013) points to, "long-term relations, frequent interactions, and mutually beneficial economic relationships" (178) as underlying effective private regulatory initiatives. However, the role of long-term commercial relationships remains poorly understood.

One possibility is that long-term commercial relationships make the common suite of private regulatory activity—rating compliance, developing remediation plans for violations, and monitoring progress—more effective (Locke, Amengual and Mangla 2009). Longer-term commercial relationships are associated with higher levels of buyer–supplier cooperation (McMillan and Woodruff 1999). Cooperation with buyers, in turn, has been associated with greater improvement under private regulation (Hugill, Short and Toffel 2016). The perceived durability of a long-term commercial relationship may also enable suppliers to invest in remediation activity which only has a return so long as the business relationship continues.

No matter whether long-term relationships have the benefits described above, they could also make suppliers more responsive to communication linking continuation of the buyer–supplier relationship to improved social compliance. Long-term relationships may increase the credibility of messages from the sourcing department about the importance of social compliance in maintaining the business relationship. Suppliers may be generally inclined to disregard claims that social performance will influence future business as empty

threats or transitory demands. However, if the two parties have established reputations through a long-term commercial relationship, suppliers may find these messages more credible, increasing the perceived risks of ignoring them. In addition, commercial relationships may become more valuable to suppliers over time due to the trust and reputation developed through repeat interactions (Macchiavello and Morjaria 2015). If long-term relationships result in more credible communication and higher-value business relationships, suppliers may invest more in improving social performance in order to retain them.

Finally, long-term commercial relationships could instead *undermine* attempts to create links between the future of the business relationship and social performance. If suppliers have enjoyed a longstanding business relationship with a buyer who has not previously linked the business relationship to social compliance, those suppliers may ignore messages claiming that expectations have changed. These suppliers have observed over many years that the buyer has continued business relationships despite persistent social noncompliance. By contrast, new suppliers likely have more malleable understandings of what buyers expect, due to their shorter histories of interaction with the buyer. A stated change in those expectations—for example, claiming that the future business relationship depends upon supplier performance in social audits—may be more credible to these short-term suppliers whose beliefs are less entrenched and who feel less secure in their understandings of the buyer's expectations.

This study empirically addresses a series of questions that emerge from the discussion above. Does the common suite of private regulatory activities—monitoring suppliers, rating compliance, and designing and supporting remediation plans—cause compliance to improve? Does improving the alignment of sourcing activity with private regulatory efforts make private regulation more effective? Finally, how does the length of the buyer–supplier commercial relationship condition the efficacy of private regulation?

Contributions

Our argument and empirical analyses make three contributions. First, we contribute to debates about private regulation. Despite widespread scholarly interest in social and environmental issues in global supply chains for nearly two decades (Frenkel 2001; Barrientos and Smith 2007; Locke 2013; Toffel, Short and Ouellet 2015; Porteous, Rammohan and Lee 2015; Lee and Tang 2017; Caro et al. 2018; Bartley 2018; Dietz, Grabs and Estrella Chong 2019; Liu et al. 2019; Bird, Short and Toffel 2019), we still have little evidence of the causal effects of private regulatory programs. We offer the first, to our knowledge, estimates of the causal effect of a suite of common supply chain regulatory practices: rating supplier performance, supporting improvement, and communicating about possible commercial consequences for noncompliance.

Second, we contribute to scholarship on private politics. Previous research shows

how social movements target firms and influence their behavior and financial performance (Eesley and Lenox 2006; King and Soule 2007; Lenox and Eesley 2009; Ingram, Yue and Rao 2010; McDonnell and King 2013; McDonnell, King and Soule 2015). Researchers have extended the study of movement impact to non-targeted peer firms in the same industry (Yue, Rao and Ingram 2013; Briscoe, Gupta and Anner 2015; Soule, Swaminathan and Tihanyi 2014). Yet contemporary social movements go further by seeking to transform practices in networks of commercial partners, in part by proxy-targeting vulnerable firms that have large supply chains (Baron 2001; Schurman and Munro 2009). By studying the impact of private regulation on supplier firms, we can understand which corporate responses satisfy movements' demands, and which serve to enhance legitimacy but have little substantive impact (Fransen 2012; Bartley and Egels-Zandén 2015; Short, Toffel and Hugill 2016).

Third, we extend scholarly research on organizational decoupling by studying its effects on other organizations. Many studies treat decoupling as an organizational outcome and investigate the factors that produce or reduce it (Coburn 2004; Fiss and Zajac 2006; Pache and Santos 2010; 2013; Dietz, Grabs and Estrella Chong 2019; Bird, Short and Toffel 2019). The present research instead treats variation in the decoupling and alignment of organizational processes as a key explanatory variable, and studies its implications for social responsibility in other firms, connected through the supply chain.

Research Design

We study private regulation in the global supply chain of Gap Inc, an international apparel retailer whose brands include Gap, Banana Republic, Old Navy, and Athleta. In 2018, it had 3,666 stores, 135,000 employees, and USD $16.6 billion in worldwide sales. Like many leading apparel retailers, Gap Inc has been targeted by activist campaigns for the labor conditions in its supplier factories. These campaigns seek to leverage Gap Inc's position as a major apparel buyer to improve working conditions in its worldwide network of supplier factories. Its suppliers are located primarily in emerging markets that dominate the global apparel trade, such as China, India, and Vietnam. These countries generally have weak regulatory institutions of labor standards enforcement and poor respect for freedom of association.

Like many multinational firms, Gap Inc has an internal supplier responsibility program to manage labor and environmental issues in its supply chain. A supplier code of conduct defines workplace standards for its suppliers, including maximum overtime hours, the content of employment contracts, and workplace safety. The code generally stipulates that suppliers comply with both local legal requirements and standards set by Gap Inc (whichever is more stringent). The supplier responsibility department audits supplier practices and implements programs to improve compliance. Social auditors visit

each factory annually for a 1–2 day inspection, depending on factory size. They inspect the physical plant for compliance with basic health and safety standards, such as adequate fire exits and ventilation. They also review documents, including worker contracts and payroll records, comparing them to legal standards. Finally, auditors interview managers and between 5 and 50 workers. Most audits are conducted by Gap Inc's supplier responsibility staff, but in a subset of factories they are conducted by the independent Better Work program, which is jointly run by the International Labor Organization and the International Finance Corporation.[4]

The social auditing system defines more than 700 categories of violations, each of which is associated with a severity level from 'low' to 'highest.' Table 1 summarizes the most frequently detected labor violations in each level over 2010–2018. The most common violations are low-severity and deal with health and safety compliance, such as protective equipment and emergency preparedness. Audits detect more severe violations, such as excessive overtime hours and insufficient rest days, somewhat less frequently. In a small minority of factories, audits detect the most serious violations, such as verbal abuse or obstructed emergency exits.

Firms use social audits because they are the best available tool to assess factory compliance in their supply chains. However, we note that social audits are not well suited for detecting violations of certain standards, such as freedom to form a union (Anner 2012). Notwithstanding this limitation, these inspection results offer the best available data on a range of labor standards—such as overtime, contracts, benefits, and safety—in developing countries and are therefore commonly used in research (Short, Toffel and Hugill 2016; Distelhorst, Hainmueller and Locke 2017; Bird, Short and Toffel 2019).

Each labor audit produces a corrective action plan and a compliance rating. The corrective action plan details all the violations detected, outlines steps the supplier should take to correct them, and provides timelines for this process. Supplier responsibility staff then provide technical support to factories as they implement the plan and monitor their progress. For example, supplier responsibility staff might explain how to improve emergency evacuation routes or how to re-organize production to reduce excessive overtime. These are typical cooperative remediation steps, documented in previous research in other organizations (Locke, Amengual and Mangla 2009).

Audits also generate a compliance rating of each factory's social performance. Categorical compliance ratings offer a shorthand to characterize the level of concern about

---

[4]Better Work audits factories in the garment industry and seeks to improve labor conditions. In 2018 it engaged over 1,400 factories across Bangladesh, Cambodia, Haiti, Indonesia, Jordan, Nicaragua, and Vietnam. URL: http://www.betterwork.org. Where Better Work operates, buyers often rely on Better Work social audits rather than their own internal audits. In certain countries, such as Cambodia, all garment exporters must participate in Better Work. In others, such as Indonesia, factories opt in to Better Work, often at the encouragement of their buyers.

Table 1: Most frequently-detected violations in labor audits, by severity

	Low-severity violations
43%	Exit routes / emergency routes
36%	Personal protective equipment
35%	Machine/Equipment safety
29%	First aid and medical
26%	Emergency procedures and evacuation drills
	Medium-severity violations
17%	One day off in 7
13%	Worker schedules
12%	Payment of benefits
10%	Reasonable emergency leave period
8%	Employment contracts availability
	High-severity violations
18%	Payment of benefits
15%	Overtime hours
10%	Overtime and incentive rates
9%	Minimum wage requirement
5%	Payment for leave not taken
	Highest-severity violations
5%	Documentation transparency
5%	Exit routes / emergency routes
4%	Verbal abuse
3%	Access to facility, workers, and records
2%	Unauthorized subcontracting

a supplier's practices, alongside the details contained in the corrective action plan. These ratings, the lowest of which is to 'fail,' are determined by a formula.[5] Each factory begins with a full score, from which points are deducted for each violation, according to their severity and whether the violation was detected in previous audits. Before 2015 violations had two levels of severity, and suppliers could earn additional points for achieving sustainability certifications such as SA8000. In 2015, a re-scaled audit scoring system created four levels of severity (see Table 1) and stopped adding points for certifications.

To assign the compliance rating, auditors return to their office and record all detected violations in a system that computes a numeric score associated with the audit. The compliance rating is entirely determined by whether the score is greater or less than a numeric threshold. Factories scoring just above the 'failing' threshold pass. Soon thereafter, Gap Inc's staff communicates the compliance rating to factory management, which is not permitted to lobby for changes in the rating.

---

[5]Gap Inc uses colors in its social compliance ratings. The highest performers are rated green, average performance yellow, and the lowest-performing factories are rated red. For shorthand, we refer to red as a 'failing' rating.

Aligning Sourcing with Supplier Responsibility

Gap Inc offers an instructive setting to study private regulation because of changes in its organization of supplier responsibility and sourcing during our study period.[6] Prior to 2016, a failing audit rating communicated that the factory's social performance was substandard and prompted supplier responsibility staff to increase monitoring and support for implementing corrective action plans. The rating helped supplier responsibility staff to allocate their scarce time across a number of factories.

In this period, supplier responsibility and sourcing largely worked in separate organizational silos. Supplier responsibility staff were responsible for monitoring conditions in factories and supporting improvement, but they had little direct control over the buyer–supplier relationship. Occasionally, egregious social responsibility violations would lead to termination of a commercial relationship, but this occurred only in rare cases. Indeed, analyzing terminations from the supply chain, we see little difference between factories that passed and failed social compliance audits over 2010–2015 (Figure 1). The sourcing department was responsible for managing commercial relationship with suppliers. So long as no zero-tolerance social or environmental issues emerged, it was generally free to make sourcing decisions based on traditional considerations: volume, price, quality, and delivery performance.

Table 2: Actions triggered when factories failed labor compliance audits

	2010–2015	2016–2019
Failed audit signals that supplier's employment practices are unacceptable.	✓	✓
Supplier responsibility personnel increase technical support for remediating violations.	✓	✓
Sourcing personnel credibly communicate possibility of commercial consequences for failing to improve.		✓

Gap Inc dismantled these silos and created stronger organizational linkages between sourcing and supplier responsibility following two events. First, the Rana Plaza disaster—in which over one thousand people were killed in the collapse of an export factory in Bangladesh—shockingly illustrated the risks to workers in the global garment industry. It resulted in increased public scrutiny of labor issues in the industry and increased awareness among senior managers of the associated risks. Activist campaigns increased pressure on

---

[6]To understand the compliance rating system and the changing relationship between sourcing and compliance, we conducted 22 interviews with Gap Inc managers in both its corporate headquarters in San Francisco and Asia office in Hong Kong.

apparel firms to improve labor conditions in their suppliers. Media coverage discussing factory labor issues and Gap Inc appears throughout the period of our study, but the number of articles quadrupled after the disaster in 2013.[7] Second, Gap Inc appointed a new leadership team in social responsibility. The new leader of social responsibility had extensive experience in the sourcing department, which provided both a deep understanding of how sourcing worked as well as personal relationships with sourcing managers.

The new leadership of supplier responsibility set a goal of zero failing factories in the supply chain. In pursuit of this performance target, several changes improved the alignment of sourcing activity with supplier responsibility. The target was shared by both the supplier responsibility department and the sourcing department. Senior managers from each reported against this goal and their progress was shared with executives at the board level. Second, new channels of communication between supplier responsibility and sourcing staff were established. Sourcing and supplier responsibility staff met monthly to review progress among all factories that recently failed audits. When suppliers failed an audit, the relevant sourcing and social responsibility managers in local offices and world headquarters received an automatic notification: "Please be advised that the following factories in your region each have a newly submitted assessment with [a failing compliance rating]." This prompted meetings to determine a course of action.

This reorganization affected both communication with failing factories and the pattern of supplier terminations. Gap Inc announced the goal of zero failing factories to its entire supply base half-way through 2016, marking a change from the past. Senior sourcing managers began attending meetings with failing factories to raise the possibility of negative commercial consequences if they were unable to bring their labor practices into compliance. Lower level sourcing staff in frequent communication with suppliers reinforced this message through both informal conversations and formal quarterly reviews. Social responsibility managers noted in our interviews the value of this new communication from sourcing personnel, reinforcing their message to suppliers about the need to improve.

Termination of failing suppliers increased but was not automatic. Suppliers usually received at least one audit cycle to remediate violations and achieve compliance. This was described by interviewees as an informal 'two strikes' rule: suppliers that failed one audit would receive at least one more chance to achieve compliance. If they failed again, termination became a possibility. Even after two failures, termination depended on case-by-case considerations and was only used in a minority of cases. Figure 1 shows that termination of failing factories increased in 2017 and 2018, especially for those that failed two audits in a row. However, 36 of the 48 cases (75%) of two or more consecutive failed

---

[7] We used the LexisNexis company filter to find all articles between 1993 and 2018 that mentioned Gap Inc and contained at least one of the following keywords: sweatshop, child labor, factory safety, factory fire, labor standards, or working conditions.

Figure 1: Factory terminations by social audit performance
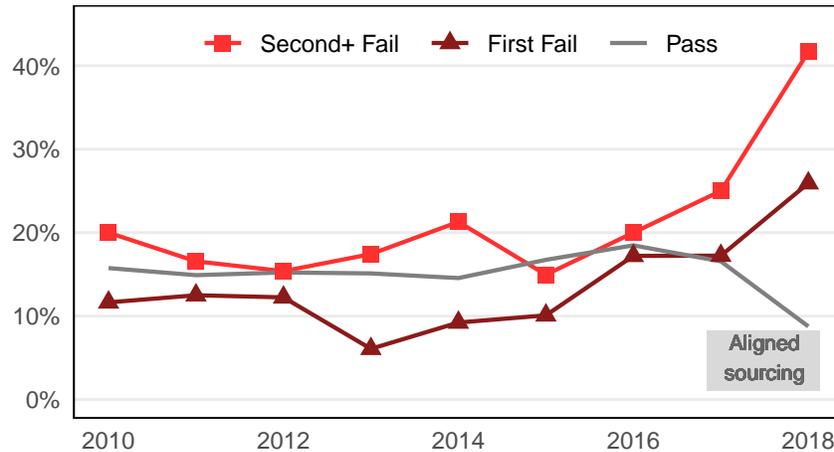
Notes. Conditional probability of factory termination within 12 months by audit rating. Factories passing, failing for the first time, and failing in two or more consecutive audits plotted separately. Results from the first two months of 2019 pooled into the 2018 data.

audits in 2017 did not lead to terminations within one year. If suppliers improved, they never observed whether the threat of termination would have been carried out.

Less than a year after these organizational changes, Gap Inc made another change to its supplier responsibility program. The interval between audits for failing factories was reduced from one year to six months. The change was prompted by the view that six months offered sufficient time to remediate even major violations. In addition, the standard yearlong interval between social audits generated uncertainty for both sourcing staff and the suppliers. This change affected only audits managed by Gap Inc, not the independent annual audits conducted by the ILO/IFC Better Work program.

Regression discontinuity design

We estimate the effects of failing a labor compliance audit under decoupled and aligned sourcing practices. Due to a rescaling of the compliance rating system one year before the changes to align sourcing with supplier responsibility (2015), our main analysis covers three periods: under the old compliance rating system (2010–2014), under the current system but *before* aligning sourcing and supplier responsibility (2015), and *after* alignment with sourcing (2016–2019). Our design uses repeated compliance audits at the same factory: a "previous" audit that fixes the categorical compliance rating (pass or fail), and a second "outcome" audit that provides our dependent variable. Over 2010–2019, we observe before-and-after audit pairs in 1,366 unique factories across 33 countries.

In addition to factory social audit results, we collected administrative records of

various factory features relevant to social compliance. These include total workers, location, previous audit scores, the duration of each factory's supplier relationship to Gap Inc, and units purchased from the factory, among others. Descriptive statistics for supplier factories under the old and current ratings systems appear in Tables 3 and 4.[8] We observe a wider range of factory covariates in the current period than under the older system. The most recent audit in our data took place on March 5, 2019.

We estimate the effect of failing a labor audit on subsequent compliance using a regression discontinuity design (Cattaneo, Idrobo and Titiunik 2018). To recover an unbiased estimate of the causal effect, the potential outcomes of the compliance score in the outcome audit must be a continuous function of the running variable (the previous audit compliance score) across the threshold of failure. By modeling these functions and computing the difference in predicted values at the threshold, we estimate a local average treatment effect of a failing rating—and the attendant actions summarized in Table 2—on compliance.[9]

If either factories or auditors adjusted scores to slightly exceed or undercut the threshold of failure, this manipulation would bias our estimates if it was correlated with potential outcomes. The compliance score calculation process allays this concern somewhat. Factories cannot lobby to change their scores, and auditors often cannot tell whether 'close call' factories scored above or below the threshold of failure until returning to their offices, inputting all violations, and receiving the final score from the compliance system. The system calculates the final score based on which violations were detected in this audit, applying additional deduction multipliers for repeat offenses.[10]

Figure 2 shows the distribution of previous audit scores near the threshold across

---

[8]A few features of the summary statistics require explanation. In some cases the length of the commercial relationship is *negative* at the time of the audit. Labor audits may occur before the factory is formally approved as a supplier. Also, some audits take place years apart rather than annually, likely due to the deactivation and later reactivation of a supplier. Finally, two covariates have been standardized to avoid revealing sensitive corporate information: units shipped from the factory and the estimated share of factory output sold to Gap Inc.

[9]Regression discontinuity designs estimate *local* causal effects, meaning the effect we estimate only pertains to factories precisely at the threshold of failure. The regression discontinuity estimate is not informative of factories whose scores fall farther from that threshold. We might imagine that factories that fail with scores farther from the threshold of failure could either: (a) feel more pressure to improve, resulting in larger effects of a failing, or (b) view achieving compliance as impossible and put less effort into improvement. It is difficult to say whether the causal effect of failure would be greater or smaller away from this threshold.

[10]Interviews with Gap Inc social auditors and managers, both in corporate headquarters and the Hong Kong office, confirm that neither auditors nor factory managers can calculate precise audit scores while the audit is in progress. Interviewees also believed factories would be unable to manipulate scores to just clear the threshold of failing audits. This would require knowing which violations would be detected by auditors, knowing the severity of those violations, and whether each represented a repeat offense (which introduces penalty multipliers). Factory managers do not have access to this full set of information.

Table 3: Descriptive statistics: old rating system (2010–2014)

	median	mean	sd	min	max	N
Previous audits						
Audit score	5	1.5	11.9	-114	14	2,366
Low-severity violations	16	19.8	15.3	0	110	2,366
Medium-severity	2	3.0	2.8	0	21	2,366
High-severity	2	2.2	2.2	0	15	2,366
Highest-severity	0	0.6	1.0	0	8	2,366
Better Work audit?	0	0.1	0.2	0	1	2,366
Days between audits	336	338.4	139.3	91	1882	2,366
Outcome audits						
Audit score	6	3.6	11.6	-57	35	2,366
Low-severity violations	21	26.1	18.0	0	115	2,366
Medium-severity	3	4.0	3.3	0	24	2,366
High-severity	2	2.9	2.6	0	19	2,366
Highest-severity	0	0.9	1.2	0	8	2,366
Better Work audit?	0	0.1	0.2	0	1	2,366
Factory characteristics						
Pre-previous audit score	6	3.1	12.8	-114	33	1,586
Relationship (years)	5	5.3	4.2	-3	18	2,277
ln(units shipped) (std)	0	-0.0	1.0	-1	1	1,133
Cust. share of output (std)	0	0.0	1.0	-1	2	1,306
Workers	708	1237.7	1735.6	1	20800	2,362
Female workers (%)	67	61.6	24.6	0	100	1,352
Manufacturer?	1	0.8	0.4	0	1	2,366
Factory in China	0	0.3	0.5	0	1	2,366
India	0	0.2	0.4	0	1	2,366
Indonesia	0	0.0	0.2	0	1	2,366
Vietnam	0	0.1	0.3	0	1	2,366
other country	0	0.3	0.5	0	1	2,366

Notes. Descriptive statistics for data under the old compliance rating system (942 total factories, unit of analysis is one audit). Audit scores zeroed at the threshold of failure; positive scores indicate the factory passed the audit. In this audit-level data factories may appear repeatedly, depending on how many total audits they experienced. For consistency with the current scoring system, we mapped violations in the old system on to the four levels of severity used today. The old system only recognized two severity levels.

Table 4: Descriptive statistics: current rating system (2015–2019)

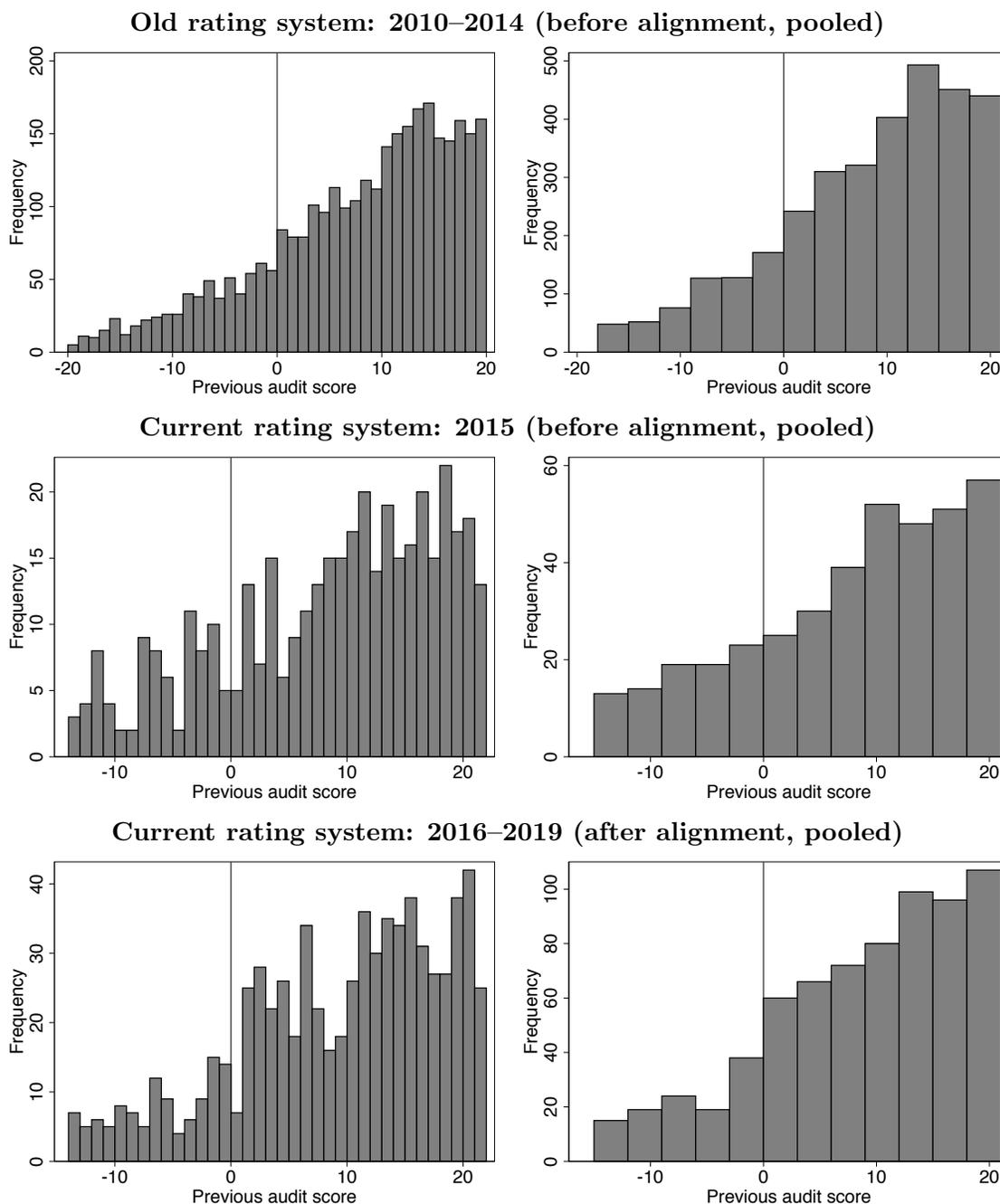|	median	mean	sd	min	max	N
Previous audits						
Audit score	8	4.4	15.7	-64	22	1,266
Low-severity violations	10	11.2	5.9	0	45	1,254
Medium-severity	2	2.3	1.7	0	12	1,254
High-severity	1	1.0	1.1	0	6	1,254
Highest-severity	0	0.3	0.6	0	3	1,254
Better work audit?	0	0.2	0.4	0	1	1,266
Days between audits	350	336.5	99.0	91	1091	1,266
Outcome audits						
Audit score	12	9.9	15.7	-64	36	1,266
Low-severity violations	9	10.3	5.8	0	44	1,112
Medium-severity	2	2.2	1.7	0	14	1,112
High-severity	1	0.9	1.0	0	7	1,112
Highest-severity	0	0.2	0.5	0	3	1,112
Better work audit?	0	0.2	0.4	0	1	1,266
Factory characteristics						
Pre-previous audit score	12	7.5	18.7	-64	36	709
Relationship (years)	6	6.9	5.3	-2	19	1,244
ln(units shipped) (std)	1	0.1	1.0	-1	2	1,266
Customer share of output (std)	-0	0.1	1.0	-1	2	1,078
Workers	807	1351.6	1695.2	1	15500	1,237
Female workers (%)	66	62.0	23.6	0	100	1,137
Manufacturer?	1	0.8	0.4	0	1	1,266
Americas	0	0.0	0.2	0	1	1,266
Mediterranean	0	0.0	0.2	0	1	1,266
North Asia	0	0.3	0.4	0	1	1,266
South Asia	0	0.3	0.5	0	1	1,266
Southeast Asia	0	0.3	0.5	0	1	1,266
Factory in China	0	0.3	0.4	0	1	1,266
India	0	0.2	0.4	0	1	1,266
Indonesia	0	0.1	0.3	0	1	1,266
Vietnam	0	0.2	0.4	0	1	1,266
other country	0	0.2	0.4	0	1	1,266

Notes. Descriptive statistics for data in the current compliance rating system (702 total factories, unit of analysis is an audit). Audit scores zeroed at the threshold of failure; positive scores indicate the factory passed the audit. Because this is audit-level data, factories may appear repeatedly, depending on how many audits they experienced.

Figure 2: Previous audit score density around the threshold of failure, pooled

Old rating system: 2010–2014 (before alignment, pooled)

Current rating system: 2015 (before alignment, pooled)

Current rating system: 2016–2019 (after alignment, pooled)

Notes. Histograms of audit scores pooling all audits within each time period. Unit of analysis is the audit, and individual factories may appear multiple times. Vertical lines show the threshold between failing and passing audit scores. Left-hand plots shows integer-wide bins and right-hand plots shows bins that are three points wide. Manipulation tests using local polynomial density estimation yield p-values of 0.48, 0.98, and 0.03, respectively. The density discontinuity in the pooled, post-alignment sample leads us to adopt a different estimation sample (Figure 3).

Figure 3: Previous audit score density post-alignment, first audits only

Current rating system: 2016–2019 (after alignment, first audits only)
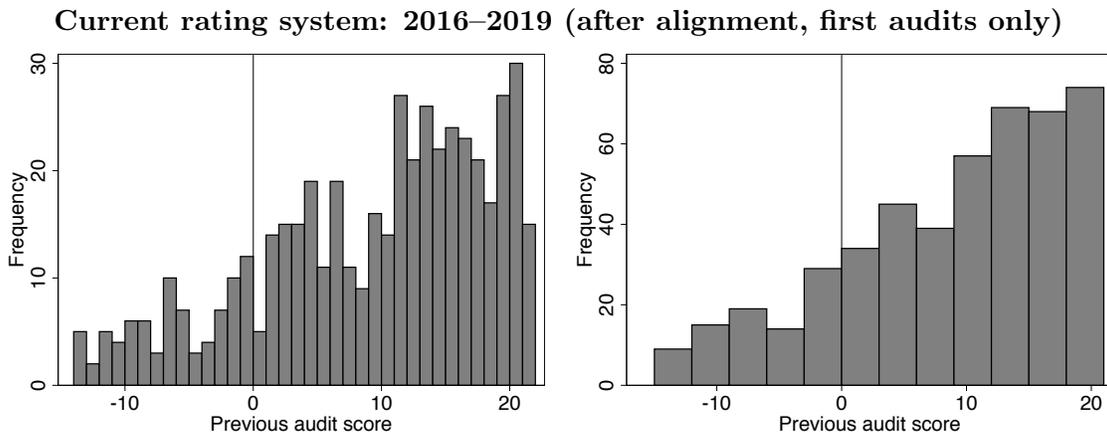
Notes. Histogram of *first* audit scores after alignment with sourcing. Each factory appears only once in this data, which serves as our running variable in the post-alignment period. Vertical lines show the threshold between failing and passing scores. Left-hand plot shows integer-wide bins and right-hand plot shows bins that are three points wide. A manipulation test using local polynomial density estimation fails to reject the null that the density of the running variable is continuous at the threshold ($p = 0.70$)

the three periods of our study. These plots pool repeated audits of the same factories, so "outcome" audits will also appear as "previous" audits when factories are audited three or more times. Observation density varies smoothly across the threshold of failure in the first two periods, and tests fail to reject the null of no difference across the threshold. However, in the third period we observe a "bulge" between two and three points above the threshold of failure.

Because the histograms in Figure 2 pool repeated audits of the same factory, one possible explanation for the above-threshold bulge is that failing factories invested in improving their compliance after alignment with sourcing (but not before). If some failing factories guessed roughly how much they needed to improve in order to pass the following audit, that could lead to an above-threshold bulge in the pooled data. If so, we expect to only observe the bulge starting in factories' *second* audits after the alignment with sourcing. To examine this, Figure 3 shows the distribution of audit scores in only the *first* audit each factory experienced post-alignment, which shows no difference in density across the threshold ($p = 0.70$). We therefore use this sample for our regression discontinuity estimates after alignment with sourcing. Each factory appears just once in this sample, and the running variable is its first post-alignment audit.[11]

---

[11]Note that using the pooled post-alignment audits yields stronger estimated effects than those reported below; doing so simply renders the assumptions required for causal identification less credible.

If potential outcomes are continuous functions across the threshold, we expect pretreatment covariates to behave similarly, exhibiting little difference on either side of the threshold. Table 5 reports regression discontinuity estimates[12] for pretreatment covariates in the post-alignment period. Across 24 statistical tests, just one rises to conventional levels of statistical significance: days elapsed between audits.[13] This was expected, as we know that audit cadence was shortened for failing factories after alignment with sourcing. In observational analyses of audit data, we find that longer time between audits is generally associated with *higher* scores, consistent with a model in which factories benefit from having more time to bring themselves into compliance. We therefore do not believe the shorter audit cycle conferred any strong advantage to the failing factories. In other features that might pose threats to inference—including the share of factory output dedicated to Gap Inc ($p =.33$), the length of the buyer–supplier relationship ($p =.68$), the preceding audit score ($p =.83$), and audits by the independent Better Work program ($p =1.00$)—we fail to detect differences between passing and failing factories at the threshold of failure.

Because each pretreatment covariate exercise generates a full page of text, we report the analyses prior to the alignment of sourcing with supplier responsibility in Appendix Tables A1 and A2. We observe somewhat fewer factory covariates in the 2010–2014 compliance rating data, but again detect no serious threats to inference in the pretreatment covariate analysis. We note that Vietnamese factories are slightly overrepresented just above the threshold of failure in the 2010–2014 data. In the current rating system before alignment (2015), we detect one imbalance in pretreatment covariates—the count of workers in failing factories was lower ($p = 0.04$). However, our analyses find a very weak, *negative* association between total workers and compliance scores; increasing the workforce by 50% is associated with a small 0.7 point (0.039 standard deviation) reduction in compliance score. Because the correlation is negative, we expect this difference would only generate a very small upward-bias in the estimated effect of failing. This imbalance is therefore unlikely to mask any positive effect of failing in the pre-alignment period.

---

[12]Throughout our regression discontinuity analyses, we use first order polynomials, triangular kernel weights, and algorithmically-selected bandwidths, following recommended procedures in Cattaneo, Idrobo and Titiunik (2018).

[13]The days elapsed between the previous and outcome audit and whether the outcome audit is conducted by Better Work are technically both "placebo outcomes." They occur after the assignment of the previous compliance rating (i.e. post-treatment).

Table 5: Continuity of pretreatment covariates - after alignment (2016-2019)

	Regression discontinuity est.			MSE-optimal	Eff.
	Est.	p-val.	95% CI	bandwidth	obs.
Low-severity violations	-0.09	0.96	[3.1, -3.3]	13.5	272
Medium-severity	0.29	0.65	[1.5, -0.9]	13.5	302
High-severity	-0.48	0.20	[0.3, -1.2]	10.7	218
Highest-severity	0.16	0.31	[0.5, -0.1]	13.8	302
Better work audit?	0.00	0.98	[0.3, -0.3]	10.3	202
Pre-previous audit score	2.13	0.83	[21.7, -17.4]	10.3	156
Subsequent audit by BW	-0.00	1.00	[0.3, -0.3]	11.8	252
Days between audits	-62.02	0.03	[-5.6, -118.4]	17.6	409
ln(units shipped) (std)	0.08	0.80	[0.7, -0.5]	15.6	360
Relationship (years)	-0.70	0.68	[2.6, -4.0]	13.0	267
Cust. output share (std)	-0.33	0.33	[0.3, -1.0]	10.1	167
Workers	-168.65	0.70	[686.4, -1023.7]	14.6	316
Female workers (%)	2.01	0.79	[16.8, -12.8]	14.9	287
Manufacturer?	-0.11	0.41	[0.2, -0.4]	8.9	180
Americas	-0.11	0.14	[0.0, -0.2]	7.8	165
Mediterranean	0.04	0.57	[0.2, -0.1]	14.1	306
North Asia	0.15	0.38	[0.5, -0.2]	8.0	165
South Asia	-0.25	0.15	[0.1, -0.6]	10.4	202
Southeast Asia	0.16	0.25	[0.4, -0.1]	13.0	275
Factory in China	0.15	0.38	[0.5, -0.2]	7.9	165
India	-0.23	0.19	[0.1, -0.6]	9.9	202
Indonesia	0.09	0.36	[0.3, -0.1]	10.8	220
Vietnam	0.06	0.62	[0.3, -0.2]	10.4	202
other country	-0.07	0.67	[0.2, -0.4]	9.3	180

Notes. Tests of continuity of pre-treatment covariates at the threshold between failing and passing audit scores. Positive values indicate that passing factories have higher values than failing ones. Estimations use first-order polynomials and algorithmically selected symmetrical bandwidths that minimize mean squared error. Estimated p-values and confidence intervals come from robust estimators described in Cattaneo, Idrobo, and Titiunik (2018). We make no adjustment for multiple comparisons.

Results

Before alignment with sourcing

Before alignment of sourcing with the supplier responsibility program—when failing signaled buyer displeasure with the factory and prompted greater attention from supplier responsibility staff—the effect of a failing rating on future compliance is near-zero. We analyze several years of data from the old compliance rating system and one year of data from the current system. Results from the older rating system appear in Figure 4, with details and four alternative algorithmic approaches to selecting the bandwidth around the threshold reported in Table 6. Each bandwidth selection technique shows a precisely estimated null effect of failing a social responsibility audit. The widest confidence interval across our four estimations spans 6.3 points, less than one half of a standard deviation in outcome compliance scores.

Figure 4: Effect before alignment with sourcing (2010–2014)
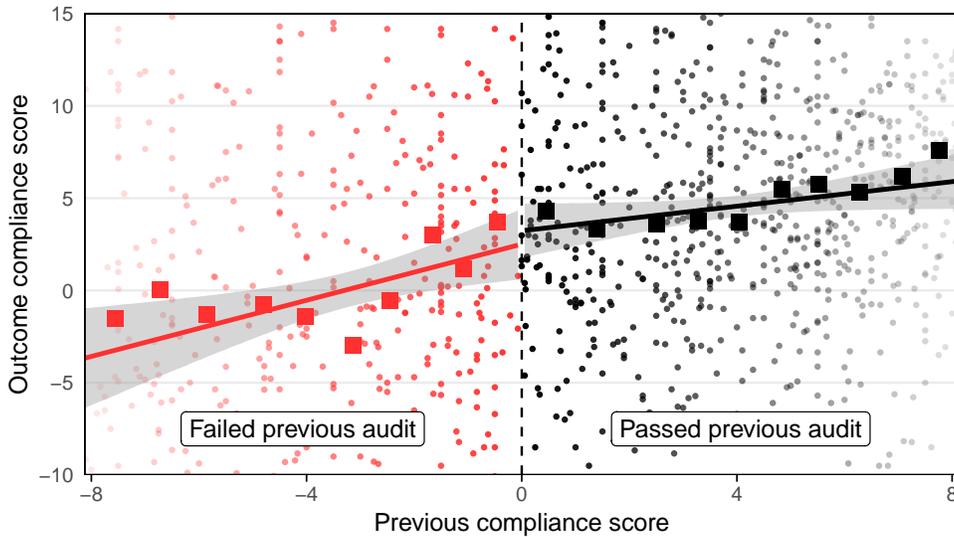
Notes. Regression discontinuity estimate of the effect of failing the previous labor compliance audit on the next compliance score *before* the alignment of sourcing and supplier responsibility (old audit scoring system, 2010–2014). Estimate uses MSE-minimizing symmetric bandwidths and triangular kernel weights. Squares show binned means of equally-sized subgroups around the discontinuity. Points show individual observations, using fading to indicate the kernel weight assigned to each observation. Details including alternative bandwidth selection techniques reported in Table 6.

Table 6: Effects before alignment with sourcing (2010-2014)

	Bandwidth properties			Regression discontinuity est.		
Procedure	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	8.1	8.1	1,164	-1.17	0.41	[-4.0, 1.6]
MSE separate	11.3	6.4	1,079	-1.26	0.34	[-3.8, 1.3]
CER symmetrical	5.6	5.6	800	-0.53	0.74	[-3.7, 2.6]
CER separate	7.8	4.4	758	-1.70	0.25	[-4.6, 1.2]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit scores in the old rating system (2010–2014). Reports four alternative data-driven approaches to selecting the bandwidth. The first approach selects a bandwidth that minimizes the expected mean squared error (MSE) of the point estimator. The second seeks to minimize the coverage error (CER) of the confidence interval around the point estimate. For each, we report both symmetrical bandwidths and separately chosen bandwidths for distributions above and below the threshold. Details of each procedure in Cattaneo, Idrobo, and Titiunik (2018).

Table 7: Effects before alignment with sourcing (2015)

	Bandwidth properties			Regression discontinuity est.		
Procedure	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	10.4	10.4	172	-2.23	0.80	[-19.6, 15.1]
MSE separate	13.1	11.7	225	-0.11	0.99	[-15.7, 15.5]
CER symmetrical	7.5	7.5	116	2.00	0.85	[-19.1, 23.1]
CER separate	9.4	8.4	140	1.17	0.90	[-17.1, 19.4]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit score. Details of each bandwidth selection procedure in Cattaneo, Idrobo, and Titiunik (2018).

Under the current compliance rating system, we again estimate near-zero effects of failing a labor compliance audit prior to alignment (2015), although these estimates have wider confidence intervals due to sparse data (Table 7). When sourcing and supplier responsibility operated in separate silos, we find no evidence that failing an audit prompted improvement.

After alignment with sourcing

A new pattern appears after the alignment of sourcing with supplier responsibility in mid-2016. Failing a previous audit caused compliance to improve by 11.6 points $[2.8, 20.4]$ (Table 8). The local average treatment effect is an increase in mean compliance score of 0.8 standard deviations, causing factories to be 22 percentage points more likely to pass their next audit. The larger squares in Figure 5 show mean compliance scores for equally-sized groups of observations above and below the threshold, illustrating a non-parametric approach to the same analysis. Mean compliance in each of the first four subgroups of failing factories below the threshold is greater than any of the first four groups of passing factories above the threshold. In short, once failing ratings triggered communication from the sourcing department about the future of the commercial relationship, failing caused supplier factories to improve.[14]

Table 8: Effects after alignment with sourcing (2016–2019)

Procedure	Bandwidth properties			Regression discontinuity est.		
	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	12.6	12.6	275	11.61	0.01	[2.8, 20.4]
MSE separate	10.5	11.5	220	12.27	0.01	[2.9, 21.6]
CER symmetrical	9.0	9.0	180	11.71	0.02	[2.1, 21.3]
CER separate	7.6	8.3	165	10.63	0.04	[0.7, 20.6]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit score. Details of each bandwidth selection procedure in Cattaneo, Idrobo, and Titiunik (2018).

Results from independent audits

Relying on in-house social audits raises some concerns about potential sources of bias. Although interviews suggested that social auditors enjoyed independence within the orga-

---

[14]Simpler analytic approaches reach similar conclusions to those from the regression discontinuity design. Examining cross-tabulations under the current rating system, prior to alignment 64% of failing factories passed their next audit. After alignment, 85% did.

Figure 5: Effect after alignment with sourcing (2016–2019)
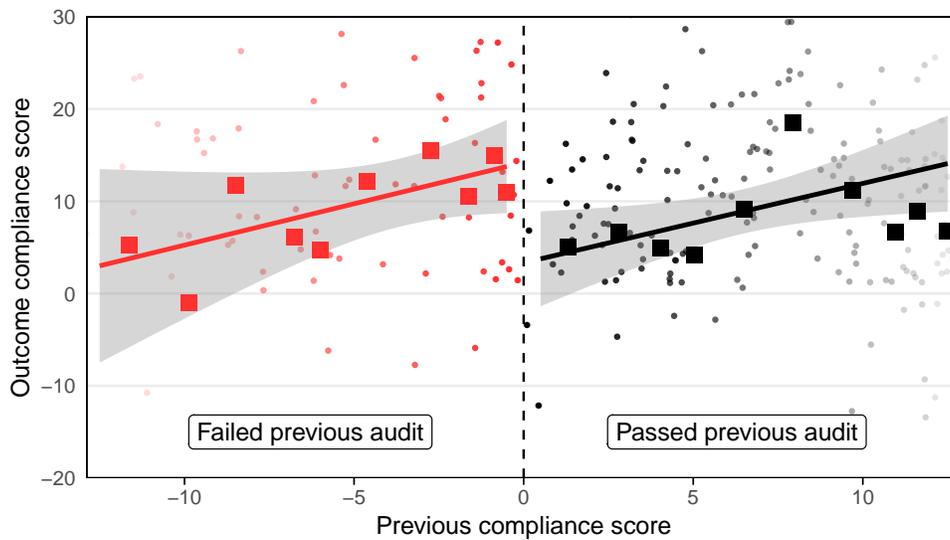
Notes. Regression discontinuity estimate of the effect of failing the previous labor compliance audit on the next compliance score *after* the alignment of sourcing with supplier responsibility (2016–2019). Estimate uses MSE-minimizing symmetric bandwidths and triangular kernel weights. Squares show binned means of equally-sized subgroups around the discontinuity. Points show individual observations, using fading to indicate their kernel weights. Details including alternative bandwidth selection techniques reported in Table 8.

nization and understood their mission to be objectively assessing the risks associated with each supplier, supplier responsibility staff were also aware that repeated failures could imperil commercial relationships. This may have subtly influenced their evaluations. In addition, although we do not believe shorter audit cadence itself confers any advantage to failing factories, the buyer's control over audit timing could bias the audit results of failing factories in unpredictable ways.

We therefore replicate the analyses above using an independent source of labor compliance information: audits conducted by the ILO/IFC Better Work program. Better Work is a regulatory program managed by the International Labour Organization and the World Bank's International Finance Corporation in Bangladesh, Cambodia, Haiti, Indonesia, Jordan, Lesotho, Nicaragua, and Vietnam. Its labor audits are conducted independently and have several features that lend greater credibility to our analysis. Better Work is overseen by advisory committees that include representatives from trade unions, employer associations, chambers of commerce, and governments. Under this governance structure, Better Work auditors operate independently from the retailers, such as Gap Inc, that source from these factories. Better Work also separates audits from factory improvement work; the ILO officials who evaluate a factory are not involved in helping that factory remediate violations. All audits are unannounced and conducted by two ILO officials. They occur annually and their timing is unrelated to previous audit scores. These features—organizational independence, program design to reduce auditor conflict of interest, and consistent audit timing—address some concerns about analyses of in-house audits.

Our study includes labor audit results from the Better Work program for 354 factories, and roughly one quarter of all factory audits since alignment with sourcing have been conducted by Better Work. We conduct a similar regression discontinuity analysis, with the cautionary note that a smaller sample size leaves us with less statistical precision. Observation density around the threshold is balanced (Appendix Figure A1), and we detect only two marginally significant covariate imbalances out of 11 statistical tests (Appendix Table A3). Factories just above the threshold have a slightly different distribution of violations, with more medium-severity violations but fewer high-severity violations.

Effects estimated from the Better Work audits appear in Table 9. Failing is estimated to cause subsequent audit scores to increase by 29 points. Due to fewer observations, these estimates are imprecise, with confidence intervals that span nearly 50 points, but all are significant at conventional levels. This result offers additional confidence in the effects of private regulation we observe after alignment with sourcing. No effect is observed in the Better Work audits prior to alignment with sourcing (Appendix Table A4).

Table 9: Effects after alignment, ILO/IFC Better Work audits only

Procedure	Bandwidth properties			Regression discontinuity est.		
	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	11.9	11.9	38	29.31	0.02	[5.5, 53.1]
MSE separate	15.6	12.5	40	25.67	0.03	[2.0, 49.3]
CER symmetrical	9.2	9.2	26	29.77	0.01	[6.4, 53.2]
CER separate	12.1	9.7	32	27.54	0.02	[4.1, 51.0]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit score after the 2016 policy change, examining only the audits conducted by the International Labor Organization / International Finance Corporation Better Work program. Details of each bandwidth selection procedure in Cattaneo, Idrobo, and Titiunik (2018).

Effects by length of buyer–supplier relationship

Does the length of the buyer–supplier relationship condition the efficacy of private regulation? One possibility is that the null average effect observed before the alignment of sourcing with supplier responsibility masked an effect that only obtained among the subgroup of suppliers with longer-term commercial relationships. After the alignment of sourcing, we may observe the effect primarily on newer suppliers (those with less experience of failing audits with few consequences) or longer-term suppliers (those that might find communications from the buyer more credible or the prospect of losing that buyer more costly).

We therefore subset our data to analyze heterogeneous effects by the length of the buyer–supplier relationship.[15] Table 10 reports effects before aligning sourcing with supplier responsibility. Suppliers are divided into two roughly equally-sized groups based on the median relationship length in the 2010–2014 data (4.5 years). We again estimate two relatively precise near-zero effects, with no significant difference based on the length of the commercial relationship.

The pattern changes after sourcing was aligned with supplier responsibility. Table 11 reports effects estimated in suppliers subsetted into equally sized groups according to the median relationship length in the 2016–2019 data (6 years). Failing caused marked improvement among suppliers with commercial relationships longer than six years. By contrast, our estimates for factories with shorter commercial relationships are nearer to

---

[15]Validation exercises for these analyses appear in the Supplementary Information.

Table 10: Effects by relationship length *before* alignment with sourcing

Procedure	Bandwidth properties			Regression discontinuity est.		
	lower	upper	eff. N	Est.	p-val.	95% CI
Supplier for under 4.5 years						
MSE symmetrical	9.0	9.0	621	-0.38	0.85	[-4.2, 3.5]
MSE separate	12.9	10.1	720	0.23	0.89	[-3.1, 3.6]
CER symmetrical	6.4	6.4	427	0.38	0.86	[-3.8, 4.6]
CER separate	9.1	7.1	519	-0.44	0.81	[-4.1, 3.2]
Supplier for over 4.5 years						
MSE symmetrical	9.1	9.1	654	-1.59	0.36	[-5.0, 1.8]
MSE separate	13.7	7.9	627	-1.81	0.24	[-4.8, 1.2]
CER symmetrical	6.6	6.6	467	-1.69	0.38	[-5.5, 2.1]
CER separate	9.9	5.7	472	-2.19	0.19	[-5.5, 1.1]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit score (2010–2014). Effects reported by equally sized factory subgroups based on shorter and longer commercial relationships with the buyer.

Table 11: Effects by relationship length *after* alignment with sourcing

Procedure	Bandwidth properties			Regression discontinuity est.		
	lower	upper	eff. N	Est.	p-val.	95% CI
Supplier for under 6 years						
MSE symmetrical	8.8	8.8	84	3.80	0.61	[-10.8, 18.4]
MSE separate	9.7	10.9	100	5.97	0.39	[-7.8, 19.7]
CER symmetrical	6.6	6.6	73	-1.64	0.82	[-16.0, 12.7]
CER separate	7.2	8.1	80	2.63	0.71	[-11.0, 16.3]
Supplier for 6+ years						
MSE symmetrical	7.3	7.3	73	25.91	0.00	[9.6, 42.3]
MSE separate	16.7	5.6	90	28.38	0.00	[10.4, 46.3]
CER symmetrical	5.4	5.4	50	20.53	0.03	[2.3, 38.8]
CER separate	12.4	4.2	64	27.03	0.01	[7.6, 46.5]

Notes. This table estimates separate regression discontinuity effects among factories that have longer and shorter commercial relationships to the buyer.

zero, although somewhat imprecisely estimated.[16] This suggests that combining longer commercial relationships with the alignment of sourcing offered the most favorable conditions for effective private regulation.

Increasing supply chain compliance through improvement or selection?

Supply chain compliance can improve through both supplier improvement and supplier terminations. The preceding analysis suggested that—once sourcing was better aligned with supplier responsibility—failing ratings caused factories to improve social compliance. However, the same effect could also be generated through selective termination near the threshold of failure. If the buyer had private information about which factories were capable of sustaining compliance, it could use that information to terminate factories in ways that bias our estimated effects of failing an audit. Specifically, if Gap Inc terminated the factories that managers knew were unable to improve and did so *more often when factories failed their first audit*, this could create higher expected potential outcomes just below the threshold of failure.[17]

We investigate the possibility of selective termination of failing factories below the threshold using similar regression discontinuity models in Table 12. If Gap Inc selectively terminated at higher rates below the threshold, we expect that failing should (at the threshold) increase the probability of termination. We find no evidence of this. This result is unsurprising given that our estimation sample uses only first audits and that Gap Inc usually offered an opportunity to improve after factories' first failure.

---

[16]The difference in effect magnitudes across long- and short-term relationships in the MSE symmetrical bandwidth estimate is $25.91 - 3.80 = 22.11$ points (Table 11). Assuming independence of these two subsamples, the standard error of this difference estimate is $\sqrt{\hat{\sigma}_1^2 + \hat{\sigma}_2^2} = \sqrt{7.47^2 + 8.34^2} = 11.2$ points (two-tailed test $p = .05$).

[17]Note that if supplier responsibility staff instead used that private information to selectively terminate factories near the threshold, but did so equally no matter whether a factory passed or failed its first audit, this behavior would not necessarily bias our estimates.

Table 12: Failing at threshold does not increase prob(termination)

Procedure	Bandwidth properties			Regression discontinuity est.		
	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	8.6	8.6	233	0.02	0.89	[-0.2, 0.23]
MSE separate	11.8	6.7	217	0.06	0.58	[-0.2, 0.28]
CER symmetrical	6.1	6.1	151	-0.04	0.72	[-0.3, 0.18]
CER separate	8.4	4.7	151	0.08	0.48	[-0.1, 0.29]

Notes. Regression discontuity estimates of the effect of failing ratings on the probability of not having a subsequent audit. Because audits are on a 12-month cadence, the sample is limited to audits conducted at least 365 days prior to the final audit in our data (March 6, 2019). Details of the bandwidth selection procedures are described in Cattaneo, Idrobo, and Titiunik (2018).

Discussion

The findings above suggest that private regulation has a positive impact on factory compliance when sourcing supports the goals of supplier responsibility programs. We find this improvement despite the presence of competitive pressures and institutional barriers to compliance noted in previous research. Yet the near-zero effects estimated prior to this organizational change are equally important. Before Gap Inc aligned sourcing with supplier responsibility, its private regulatory program relied primarily on communicating disapproval and increasing assistance to failing factories from supplier responsibility personnel. Previous research indicates that this approach is common (Locke 2013). Our findings suggest that core elements of private regulation—assigning social compliance ratings and increasing support for remediation—have little causal effect on supplier compliance in the absence of alignment with sourcing. These findings reinforce claims made by many social movement organizations that sourcing departments need to support the goal of socially responsible supply chains.[18]

This result has implications for research on private politics and decoupling. Broadly, research on private politics analyzes when firms concede to pressure from stakeholders and adopt policies to maintain legitimacy (Lenox and Eesley 2009; Ingram, Yue and Rao 2010; McDonnell and King 2013; McDonnell, King and Soule 2015; Yue, Rao and Ingram 2013; Briscoe, Gupta and Anner 2015; Soule, Swaminathan and Tihanyi 2014). Yet, firms may adopt standards and invest resources into social compliance without achieving the goals articulated by social movements. If this is caused by the decoupling of core

---

[18]Organizations advocating for greater alignment of sourcing (purchasing) practices with supply chain responsibility programs include the Ethical Trade Initiative, Fair Food Program, Fair Labor Association, Oxfam, and Worker Rights Consortium.

business practices from policies adopted under stakeholder duress, breaking down these organizational silos may offer a solution.

There are scope conditions to our argument that multinational buyers can make private regulation more effective by aligning the activity of sourcing departments with social responsibility goals. First, Gap Inc is a major customer of these supplier factories, often purchasing more than one-third of their output, similar to other large apparel firms (Locke 2013). Losing this particular buyer would likely impose significant costs on these suppliers. By contrast, smaller buyers purchase in smaller volumes, leading suppliers to place lower value on maintaining the business relationship in the face of demands to improve social compliance. Second, stable sourcing relationships and the expectation of future business may also be necessary conditions for the effects we observe. Half of Gap Inc's suppliers have been in a commercial relationship for more than six years, and we found that failing ratings had the clearest effect on this subgroup (Table 11). Not all supply chains are characterized by such stable relations. A study of the apparel industry in Bangladesh, one of the world's leading garment exporters, found that the average length of a buyer–supplier relationship was just two years (Cajal Grossi, Macchiavello and Noguera 2019). Gap Inc's comparatively long commercial relationships may make the promise of future business more credible. When commercial relationships are less stable, suppliers may place lower value on the future of the commercial relationship, anticipating it may end even if they invest in improving their social compliance.

Suppliers in longer-term commercial relationships were more responsive to private regulation after alignment with sourcing. This finding is consistent with the hypothesis that long-term relationships enable more credible communication between buyer and supplier, which in turn prompts these suppliers to invest more in improving compliance. However, we cannot exclude other possibilities. It is also possible that communication about the future relationship was equally credible among suppliers with longer- and shorter-term relationships, but long-term relationships enabled more effective buyer–supplier cooperation in remediating violations. The precise mechanism linking long-term relationships to improved compliance awaits future research. What our analysis shows is that long-term relationships alone were not sufficient to prompt improvement. Heterogeneous effects emerged only *after* aligning sourcing with supplier responsibility.

In addition to the private regulatory practices analyzed in this study, buyers have other tools to influence the social compliance of their suppliers. Our study has less to say about improvements prompted by supplier selection or interventions to transform suppliers' management practices (Bloom et al. 2013). For example, even before aligning sourcing with supplier responsibility, many factories with deplorable labor conditions were likely screened out by Gap Inc's social audits. They never entered the supply chain, and we therefore do not observe their data. If certain social practices are disqualifying for exporting to multinational enterprises, supplier selection may generate incentives for labor upgrading in

Table 13: How much do failing factories improve?

	Failed audit	Next audit	diff	std err
Total violations	21.6	16.3	-5.2	0.85
Counts by severity				
Low-severity violations	13.5	9.8	-3.6	0.66
Medium-severity	3.3	2.4	-0.9	0.20
High-severity	1.6	1.1	-0.5	0.14
Highest-severity	0.8	0.2	-0.5	0.07

Notes. Change in violations after receiving a failing rating, across all failing audits after alignment with sourcing. In *t*-tests for differences in means assuming unequal variances, $p < .001$ for all comparisons. $N = 130$.

order to gain access to export markets, as suggested by survey evidence from Vietnamese exporters reported in Malesky and Mosley (2018). Other interventions have also been shown to improve factory labor standards. Recent research suggests that interventions by multinational enterprises to improve worker voice (Boudreau 2018), upgrade manufacturing management practices (Distelhorst, Hainmueller and Locke 2017), or optimize the wage and bonus system (O'Rourke and Lollo 2018) offer alternative pathways to improving working conditions in global supply chains.

Finally, although we find that private regulation can improve labor compliance of suppliers, stronger actions are likely needed to bring suppliers into *full* compliance with international standards. Table 13 shows average audit-to-audit improvement in all failing factories since the organizational change to align sourcing with supplier responsibility. Failing factories markedly improved in their subsequent audits, reducing highest-severity violations by 75% (from 0.8 to 0.2 on average). Yet subsequent audits of failing factories still detected an average of 16 violations and more than one high-severity violation. It is possible that this is merely one stop on the way to even higher levels of compliance, but it seems more likely that suppliers will maintain the minimal level of compliance needed to minimize the risk of termination. Future improvement may depend on supply chain buyers adopting increasingly stringent definitions of social compliance "failure."

References

Amengual, Matthew. 2010. "Complementary Labor Regulation: The Uncoordinated Combination of State and Private Regulators in the Dominican Republic." *World Development* 38(3):405–414.

Amengual, Matthew, Greg Distelhorst and Danny Tobin. 2019. "Global Purchasing as Labor Regulation: The Missing Middle." *Industrial and Labor Relations Review, forthcoming* .

Anner, Mark. 2012. "Corporate Social Responsibility and Freedom of Association Rights The Precarious Quest for Legitimacy and Control in Global Supply Chains." *Politics & Society* 40(4):609–644.

Anner, Mark. Forthcoming. "Squeezing workers' rights in global supply chains: purchasing practices in the Bangladesh garment export sector in comparative perspective." *Review of International Political Economy* .

Anner, Mark, Jennifer Bair and Jeremy Blasi. 2013. "Toward joint liability in global supply chains: Addressing the root causes of labor violations in international subcontracting networks." *Comp. Lab. L. & Pol'y J.* 35:1.

Baron, David P. 2001. "Private politics, corporate social responsibility, and integrated strategy." *Journal of Economics & Management Strategy* 10(1):7–45.

Baron, David P. 2003. "Private politics." *Journal of Economics & Management Strategy* 12(1):31–66.

Barrientos, S. and S. Smith. 2007. "Do Workers Benefit from Ethical Trade? Assessing Codes of Labour Practice in Global Production Systems." *Third World Quarterly* 28(4):713–729.

Bartley, Tim. 2003. "Certifying forests and factories: States, social movements, and the rise of private regulation in the apparel and forest products fields." *Politics & Society* 31(3):433–464.

Bartley, Tim. 2007. "Institutional Emergence in an Era of Globalization: The Rise of Transnational Private Regulation of Labor and Environmental Conditions." *American Journal of Sociology* 113(2):297–351.

Bartley, Tim. 2018. *Rules without rights: Land, labor, and private authority in the global economy.* Oxford University Press.

Bartley, Tim and Curtis Child. 2011. "Movements, markets and fields: The effects of anti-sweatshop campaigns on US firms, 1993-2000." *Social Forces* 90(2):425–451.

Bartley, Tim and Niklas Egels-Zandén. 2015. "Beyond decoupling: unions and the leveraging of corporate social responsibility in Indonesia." *Socio-Economic Review* 14(2):231–255.

Bird, Yanhua, Jodi L Short and Michael W Toffel. 2019. "Coupling Labor Codes of Conduct and Supplier Labor Practices: The Role of Internal Structural Conditions." *Organization Science* .

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

Bloom, Nicholas and John Van Reenen. 2010. "Why do management practices differ across firms and countries?" *Journal of economic perspectives* 24(1):203–24.

Boudreau, Laura. 2018. "Multinational enforcement of labor law: Experimental evidence from Bangladesh's apparel sector." *Working Paper* .

URL: *https://goo.gl/b7CRTH*

Briscoe, Forrest, Abhinav Gupta and Mark S Anner. 2015. "Social activism and practice diffusion: How activist tactics affect non-targeted organizations." *Administrative Science Quarterly* 60(2):300–332.

Bromley, Patricia and Walter W Powell. 2012. "From smoke and mirrors to walking the talk: Decoupling in the contemporary world." *Academy of Management annals* 6(1):483–530.

Cajal Grossi, Julia, Rocco Macchiavello and Guillermo Noguera. 2019. "International buyers' sourcing and suppliers' markups in Bangladeshi garments.".

Caro, Felipe, Prashant Chintapalli, Kumar Rajaram and Chris S Tang. 2018. "Improving Supplier Compliance Through Joint and Shared Audits with Collective Penalty." *Manufacturing & Service Operations Management* 20(2):363–380.

Cashore, Benjamin William, Graeme Auld and Deanna Newsom. 2004. *Governing through markets: Forest certification and the emergence of non-state authority.* Yale University Press.

Cattaneo, Matias D, Nicolas Idrobo and Rocıo Titiunik. 2018. *A Practical Introduction to Regression Discontinuity Designs: Volume 1.* Cambridge University Press.

Coburn, Cynthia E. 2004. "Beyond decoupling: Rethinking the relationship between the institutional environment and the classroom." *Sociology of education* 77(3):211–244.

Coslovsky, Salo V and Richard Locke. 2013. "Parallel Paths to Enforcement: Private Compliance, Public Regulation, and Labor Standards in the Brazilian Sugar Sector." *Politics & Society* 41(4):497–526.

Dickson, M.A. 2018. "Better Buying Purchasing Practices Index Report Fall 2018: Purchasing Practices Performance in Apparel, Footwear, and Household Textile Supply Chains.".

Dietz, Thomas, Janina Grabs and Andrea Estrella Chong. 2019. "Mainstreamed voluntary sustainability standards and their effectiveness: Evidence from the Honduran coffee sector." *Regulation & Governance* .

Distelhorst, Greg, Jens Hainmueller and Richard M Locke. 2017. "Does lean improve labor standards? Management and social performance in the Nike supply chain." *Management Science* 63(3):707–728.

Distelhorst, Greg, Richard M Locke, Timea Pal and Hiram M Samel. 2015. "Production Goes Global, Compliance Stays Local: Private regulation in the global electronics industry." *Regulation & Governance* 9:224–242.

Eesley, Charles and Michael J Lenox. 2006. "Firm responses to secondary stakeholder action." *Strategic Management Journal* 27(8):765–781.

Elliott, Kimberly A. and Richard B. Freeman. 2003. *Can Labor Standards Improve Under Globalization?* Peterson Institute.

Fiss, Peer C and Edward J Zajac. 2006. "The symbolic management of strategic change: Sensegiving via framing and decoupling." *Academy of Management Journal* 49(6):1173–1193.

Fransen, Luc. 2012. "Multi-stakeholder governance and voluntary programme interactions: legitimation politics in the institutional design of Corporate Social Responsibility." *Socio-Economic Review* 10(1).

Frenkel, Stephen J. 2001. "Globalization, athletic footwear commodity chains and employment relations in China." *Organization studies* 22(4):531–562.

Fung, Archon, Dara O'Rourke and Charles F. Sabel. 2001. *Can We Put an End to*

Sweatshops? A New Democracy Forum on Raising Global Labor Standards. Beacon Press.

Gereffi, Gary. 1999. "International Trade and Industrial Upgrading in the Apparel Commodity Chain." *Journal of International Economics* 48(1):37–70.

Gibbons, Robert. 2005. "Incentives between firms (and within)." *Management science* 51(1):2–17.

Harrison, Ann and Jason Scorse. 2010. "Multinationals and Anti-Sweatshop Activism." *The American Economic Review* 100(1):247–273.

Helper, Susan and Rebecca Henderson. 2014. "Management practices, relational contracts, and the decline of General Motors." *Journal of Economic Perspectives* 28(1):49–72.

Hugill, Andrea R, Jodi L Short and Michael W Toffel. 2016. "Beyond Symbolic Responses to Private Politics: Examining Labor Standards Improvement in Global Supply Chains." Harvard Business School Working Paper 17-001.

ILO. 2017. "Purchasing practices and working conditions in global supply chains: Global Survey results.".

Ingram, Paul, Lori Qingyuan Yue and Hayagreeva Rao. 2010. "Trouble in store: Probes, protests, and store openings by Wal-Mart, 1998–2007." *American Journal of Sociology* 116(1):53–92.

Jira, Chonnikarn and Michael W Toffel. 2013. "Engaging supply chains in climate change." *Manufacturing & Service Operations Management* 15(4):559–577.

Kalkanci, Basak and Erica L Plambeck. 2019. "Reveal the supplier list? A trade-off in capacity vs. responsibility." *Manufacturing & Service Operations Management.* .

Kalkanci, Basak, Erjie Ang and Erica L Plambeck. 2016. Strategic disclosure of social and environmental impacts in a supply chain. In *Environmentally responsible supply chains.* Springer pp. 223–239.

King, Brayden G and Sarah A Soule. 2007. "Social movements as extra-institutional entrepreneurs: The effect of protests on stock price returns." *Administrative Science Quarterly* 52(3):413–442.

Lee, Hau L and Christopher S Tang. 2017. "Socially and environmentally responsible value chain innovations: New operations management research opportunities." *Management Science* 64(3):983–996.

Lenox, Michael J and Charles E Eesley. 2009. "Private environmental activism and the selection and response of firm targets." *Journal of Economics & Management Strategy* 18(1):45–73.

Liu, Xiaojin, Anant Mishra, Susan Goldstein and Kingshuk K Sinha. 2019. "Toward improving factory working conditions in developing countries: An empirical analysis of Bangladesh Ready-Made Garment factories." *Manufacturing & Service Operations Management* 21(2):251–477.

Locke, R and Hiram Samel. 2018. "Beyond the workplace: "upstream" business practices and labor standards in the global electronics industry." *Studies in Comparative International Development* 53(1):1–24.

Locke, Richard M. 2013. *The Promise and Limits of Private Power: Promoting Labor Standards in a Global Economy.* Cambridge University Press.

Locke, Richard M, Ben A Rissing and Timea Pal. 2013. "Complements or substitutes? Private codes, state regulation and the enforcement of labour standards in global supply chains." *British Journal of Industrial Relations* 51(3):519–552.

Locke, Richard M., Fei Qin and Alberto Brause. 2007. "Does Monitoring Improve Labor

Standards? Lessons from Nike." *Industrial and Labor Relations Review* 61(1):3–31.

Locke, Richard M., Matthew Amengual and Akshay Mangla. 2009. "Virtue out of Necessity? Compliance, Commitment, and the Improvement of Labor Conditions in Global Supply Chains." *Politics & Society* 37(3):319.

Macchiavello, Rocco and Ameet Morjaria. 2015. "The value of relationships: evidence from a supply shock to Kenyan rose exports." *American Economic Review* 105(9):2911–45.

Malesky, Edmund J and Layna Mosley. 2018. "Chains of Love? Global Production and the Firm-Level Diffusion of Labor Standards." *American Journal of Political Science* 62(3):712–728.

McDonnell, Mary-Hunter, Brayden G King and Sarah A Soule. 2015. "A dynamic process model of private politics: Activist targeting and corporate receptivity to social challenges." *American Sociological Review* 80(3):654–678.

McDonnell, Mary-Hunter and Brayden King. 2013. "Keeping up appearances: Reputational threat and impression management after social movement boycotts." *Administrative Science Quarterly* 58(3):387–419.

McMillan, John and Christopher Woodruff. 1999. "Interfirm relationships and informal credit in Vietnam." *The Quarterly Journal of Economics* 114(4):1285–1320.

Meyer, John W and Brian Rowan. 1977. "Institutionalized organizations: Formal structure as myth and ceremony." *American journal of sociology* 83(2):340–363.

O'Rourke, Dara and Niklas Lollo. 2018. "Productivity, Profits, and Pay: A Field Experiment Analyzing the Impacts of Compensation Systems in an Apparel Factory." *UC-Berkeley IRLE Working Paper 104-18* .

Oxfam. 2010. "Better Jobs in Better Supply Chains." *Briefings for Business No. 5* .

Pache, Anne-Claire and Filipe Santos. 2010. "When worlds collide: The internal dynamics of organizational responses to conflicting institutional demands." *Academy of management review* 35(3):455–476.

Pache, Anne-Claire and Filipe Santos. 2013. "Inside the hybrid organization: Selective coupling as a response to competing institutional logics." *Academy of Management Journal* 56(4):972–1001.

Porteous, Angharad, Sonali Rammohan and Hau Lee. 2015. "Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties." *Production and Operations Management* 24(9):1402–1413.

Raj-Reichert, Gale. 2013. "Safeguarding labour in distant factories: Health and safety governance in an electronics global production network." *Geoforum* 44:23–31.

Reinecke, Juliane, Jimmy Donaghey, Nancy Bocken and Lucas Lauriano. 2019. Business Models and Labour Standards: Making the Connection. Technical report Ethical Trading Initiative.

Riisgaard, L. 2009. "Global Value Chains, Labor Organization and Private Social Standards: Lessons from East African Cut Flower Industries." *World Development* 37(2):326–340.

Sako, Mari. 2004. "Supplier development at Honda, Nissan and Toyota: comparative case studies of organizational capability enhancement." *Industrial and Corporate Change* 13(2):281–308.

Schurman, Rachel and William Munro. 2009. "Targeting capital: A cultural economy approach to understanding the efficacy of two anti–genetic engineering movements." *American Journal of Sociology* 115(1):155–202.

Seidman, Gay. 2007. *Beyond the Boycott: Labor Rights, Human Rights, and Transnational*

Activism. Russell Sage Foundation Publications.

Short, Jodi L, Michael W Toffel and Andrea R Hugill. 2016. "Monitoring global supply chains." *Strategic Management Journal* 37(9):1878–1897.

Soule, Sarah A, Anand Swaminathan and Laszlo Tihanyi. 2014. "The diffusion of foreign divestment from Burma." *Strategic Management Journal* 35(7):1032–1052.

Tanaka, Mari. 2017. "Exporting Sweatshops? Evidence from Myanmar." *Review of Economics and Statistics* pp. 1–44.

Thorlakson, Tannis, Jens Hainmueller and Eric F Lambin. 2018. "Improving environmental practices in agricultural supply chains: The role of company-led standards." *Global environmental change* 48:32–42.

Thorlakson, Tannis, Joann F de Zegher and Eric F Lambin. 2018. "Companies' contribution to sustainability through global supply chains." *Proceedings of the National Academy of Sciences* pp. 2072–2077.

Toffel, Michael W, Jodi L Short and Melissa Ouellet. 2015. "Codes in context: How states, markets, and civil society shape adherence to global labor standards." *Regulation & Governance* 9(3):205–223.

Vogel, David. 2005. *The market for virtue: The potential and limits of corporate social responsibility.* Brookings Institution Press.

Vogel, David. 2010. "The private regulation of global corporate conduct: achievements and limitations." *Business & Society* 49(1):68–87.

Walker, Edward T, Andrew W Martin and John D McCarthy. 2008. "Confronting the state, the corporation, and the academy: The influence of institutional targets on social movement repertoires." *American Journal of Sociology* 114(1):35–76.

Yang, Yujeong and Mary Gallagher. 2017. "Moving In and Moving Up? Labor Conditions and China's Changing Development Model." *Public Administration and Development* 37(3):160–175.

Yue, Lori Qingyuan, Hayagreeva Rao and Paul Ingram. 2013. "Information spillovers from protests against corporations: A tale of Walmart and Target." *Administrative Science Quarterly* 58(4):669–701.

Appendix

Table A1: Continuity of pretreatment covariates before alignment (2010–2014, old rating system)

| | Regression discontinuity est. | | | MSE-optimal | Eff. |
	Est.	p-val.	95% CI	bandwidth	obs.
Low-severity violations	0.011	1.00	[4.65, -4.63]	7.83	1,127
Medium-severity	-0.645	0.12	[0.18, -1.47]	11.21	1,641
High-severity	-0.523	0.11	[0.12, -1.16]	10.73	1,566
Highest-severity	-0.005	0.98	[0.30, -0.31]	12.93	1,953
Better Work audit?	0.017	0.61	[0.08, -0.05]	12.94	1,954
Days between audits	16.481	0.43	[57.30, -24.33]	7.20	1,025
Better Work audit?	0.017	0.61	[0.08, -0.05]	12.94	1,954
Pre-previous audit score	0.800	0.61	[3.87, -2.27]	11.47	1,160
Relationship (years)	1.099	0.19	[2.75, -0.55]	7.54	1,064
ln(units shipped) (std)	0.119	0.60	[0.57, -0.33]	9.37	653
Cust. share of output (std)	0.041	0.84	[0.44, -0.36]	11.73	991
Workers	160.754	0.55	[685.66, -364.16]	11.18	1,634
Female workers (%)	2.544	0.59	[11.74, -6.65]	10.90	955
Manufacturer?	-0.039	0.50	[0.08, -0.15]	8.61	1,260
Factory in China	0.014	0.84	[0.15, -0.12]	12.72	1,918
India	-0.010	0.89	[0.13, -0.15]	10.46	1,492
Indonesia	0.011	0.75	[0.08, -0.06]	8.44	1,214
Vietnam	-0.106	0.04	[-0.00, -0.21]	7.82	1,127
other country	0.103	0.17	[0.25, -0.04]	8.98	1,297

Notes. Tests of continuity of pre-treatment covariates at the threshold of failure in the old ratings system (2010-2014). Positive values indicate that factories that passed have higher values than those that failed. Estimations use first-order polynomials and algorithmically selected symmetrical bandwidths that minimize mean squared error. Estimated p-values and confidence intervals come from robust estimators described in Cattaneo, Idrobo, and Titiunik (2018).

Table A2: Continuity of pretreatment covariates - before alignment (2015)

| | Regression discontinuity est. | | | MSE-optimal | Eff. |
	Est.	p-val.	95% CI	bandwidth	obs.
Low-severity violations	-0.69	0.71	[2.9, -4.3]	15.0	280
Medium-severity	0.17	0.77	[1.3, -1.0]	16.0	299
High-severity	-0.08	0.81	[0.6, -0.8]	12.2	219
Highest-severity	0.09	0.61	[0.5, -0.3]	17.6	340
Better work audit?	-0.14	0.25	[0.1, -0.4]	9.8	172
Pre-previous audit score	-22.03	0.28	[17.8, -61.9]	8.5	12
Subsequent audit by BW	-0.15	0.42	[0.2, -0.5]	9.7	172
Days between audits	-7.94	0.89	[101.7, -117.6]	10.2	172
ln(units shipped) (std)	-0.28	0.47	[0.5, -1.0]	11.2	193
Relationship (years)	-1.65	0.38	[2.0, -5.3]	12.2	219
Cust. output share (std)	0.06	0.89	[0.9, -0.8]	9.7	143
Workers	-1530.26	0.04	[-44.6, -3015.9]	7.9	138
Female workers (%)	12.61	0.14	[29.2, -4.0]	12.9	218
Manufacturer?	-0.14	0.33	[0.1, -0.4]	12.5	239
Americas	-0.04	0.54	[0.1, -0.2]	9.3	155
Mediterranean	-0.00	0.99	[0.1, -0.1]	7.2	116
North Asia	0.26	0.17	[0.6, -0.1]	9.0	155
South Asia	-0.30	0.10	[0.1, -0.7]	10.0	172
Southeast Asia	0.09	0.64	[0.5, -0.3]	12.0	221
Factory in China	0.26	0.17	[0.6, -0.1]	9.0	155
India	-0.26	0.13	[0.1, -0.6]	10.2	172
Indonesia	0.01	0.89	[0.1, -0.1]	11.2	193
Vietnam	0.00	0.99	[0.3, -0.3]	15.3	282
other country	-0.02	0.89	[0.3, -0.3]	10.4	172

Notes. Tests of continuity of pre-treatment covariates at the threshold between failing and passing audit scores. Positive values indicate that passing factories have higher values than failing ones. Estimations use first-order polynomials and algorithmically selected symmetrical bandwidths that minimize mean squared error. Estimated p-values and confidence intervals come from robust estimators described in Cattaneo, Idrobo, and Titiunik (2018). We make no adjustment for multiple comparisons.

Table A3: Continuity of covariates, ILO/IFC Better Work audits only

|	Regression discontinuity est.			MSE-optimal	Eff.
	Est.	p-val.	95% CI	bandwidth	obs.
Low-severity violations	-3.84	0.26	[2.8, -10.5]	13.5	40
Medium-severity	-2.95	0.08	[0.4, -6.3]	10.1	29
High-severity	3.11	0.07	[6.5, -0.2]	9.3	25
Highest-severity	-0.57	0.25	[0.4, -1.5]	10.2	29
Pre-previous audit score	14.53	0.35	[44.9, -15.9]	10.9	24
ln(units shipped) (std)	0.41	0.55	[1.7, -0.9]	10.3	30
Days between audits	-9.31	0.70	[38.4, -57.1]	9.0	26
Relationship (years)	0.08	0.99	[9.1, -9.0]	12.2	38
Cust. output share (std)	0.16	0.83	[1.6, -1.3]	17.0	65
Workers	-1074.53	0.59	[2823.5, -4972.6]	12.9	40
Female workers (%)	-2.53	0.85	[22.8, -27.9]	10.5	29
Manufacturer?	0.04	0.65	[0.2, -0.1]	6.3	17
Factory in Bangladesh	0.19	0.65	[1.0, -0.6]	11.4	32
Cambodia	0.40	0.31	[1.2, -0.4]	10.2	30
Indonesia	0.12	0.75	[0.9, -0.6]	12.2	38
Vietnam	-0.24	0.56	[0.6, -1.1]	12.0	38
other country	-0.50	0.14	[0.2, -1.1]	7.6	24

Notes. Tests of continuity of pre-treatment covariates at the threshold between failing and passing audit scores for the Better Work audits, after the 2016 policy change. Positive values indicate that passing factories have higher values than failing ones. Estimations use first-order polynomials and algorithmically selected symmetrical bandwidths that minimize mean squared error. Estimated p-values and confidence intervals come from robust estimators described in Cattaneo, Idrobo, and Titiunik (2018). We make no adjustment for multiple comparisons.

Table A4: Effects before alignment, ILO/IFC Better Work audits only (2010–2014)

	Bandwidth properties			Regression discontinuity est.		
Procedure	lower	upper	eff. N	Est.	p-val.	95% CI
MSE symmetrical	8.8	8.8	72	-4.30	0.42	[-14.8, 6.2]
MSE separate	9.5	7.3	67	-4.29	0.42	[-14.8, 6.2]
CER symmetrical	7.1	7.1	57	-2.47	0.64	[-12.9, 8.0]
CER separate	7.7	5.9	54	-2.95	0.59	[-13.7, 7.8]

Notes. Regression discontinuity estimates of the effect of failing ratings on subsequent audit score *before* alignment with sourcing (2010–2014). Exclusively uses audits conducted by the ILO/IFC Better Work program. Details of each bandwidth selection procedure in Cattaneo, Idrobo, and Titiunik (2018).

Figure A1: Audit score density around the threshold of failure: Better Work audits

Current system: 2016–2019 (after alignment, Better Work audits)
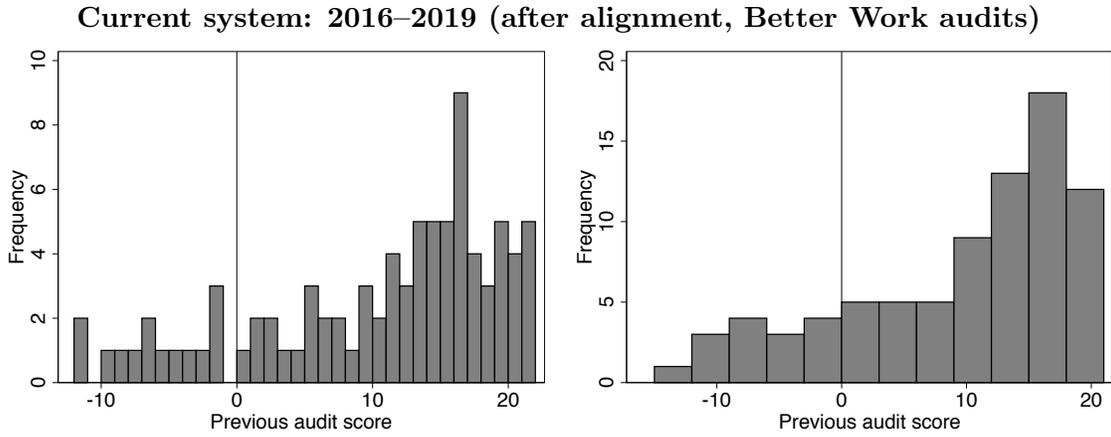
Notes. Histograms of audit scores, with a vertical line showing the threshold between failing and passing audit scores. This plot shows the first audit after alignment of sourcing with supplier responsibility, only if both the audit and the next audit were conducted by the ILO Better Work program. Left-hand plot shows all scores (each is an integer) and right-hand plot shows bins that are three points wide. A manipulation test using local polynomial density estimation fails to reject the null of no sorting around the threshold ($p = .99$).