Essays on Audit Program Design, Audit Quality, and Violation Resolution

HARVARD UNIVERSITY
Graduate School of Arts and Sciences

DISSERTATION ACCEPTANCE CERTIFICATE

The undersigned, appointed by the Committee for the PhD in Business Administration have examined a dissertation entitled

Essays on Audit Program Design, Audit Quality, and Violation Resolution

Presented by Ashley Kristin Palmarozzo

candidate for the degree of Doctor of Philosophy and hereby certify that it is worthy of acceptance.

Signature ______________________

Michael Toffel, Chair

Signature ______________________

Hong Luo

Signature ______________________

Feng Zhu

Date: April 5, 2023

Essays on Audit Program Design, Audit Quality, and Violation Resolution

A dissertation presented

by

Ashley Kristin Palmarozzo

to

Harvard Business School

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Business Administration

Harvard University

Cambridge, Massachusetts

April 5, 2023

Dissertation Advisor: Michael Toffel | Ashley Palmarozzo

Essays on Audit Program Design, Audit Quality, and Violation Resolution

Downstream businesses that utilize global suppliers frequently use auditing programs to monitor their suppliers' working conditions and are often deployed to address reputational concerns associated with procuring from unregulated suppliers. Despite their widespread use, it is unclear how audit program design decisions influence the effectiveness of those programs. This dissertation explores three features of audit programs that may affect program effectiveness.

Chapter 1 documents the audit quality effects of four different auditor sourcing strategies. Using proprietary data from a global apparel brand, we find that insourced auditors yield higher quality audits, by recording more audit violations, than those done by outsourced auditors, and that third-party audit quality increases with the use of concurrent sourcing or rotational sourcing.

Chapter 2 explores how differences in audit format can influence audit quality in the context of management system standard audits. It finds that audits are of lower quality (resulting in fewer violations) when conducted remotely than when conducted in-person, possibly because remote auditors face greater difficulty obtaining information to document violations. It also finds that the remote audit quality decrement increases with multi-auditor teams and decreases when teams have more prior in-person site visits.

Finally, chapter 3 identifies two aspects of the buyer-supplier's economic relationship that may influence the resolution of violations in supplier working conditions. It finds that the odds of violation resolution increase when a greater portion of supplier output goes to the buyer that is auditing them, and when there are more alternative suppliers for the buyer to choose from.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

The road to a dissertation can feel long and lonely. I therefore have such gratitude for the support of mentors, colleagues, family, and friends, and I have the deepest gratitude for all who have helped me along my six-year Ph.D. journey.

First, my deepest thanks to my primary advisor, Mike, who has served as a mentor from day one of my journey. Thank you for taking me on as a student and showing me how to conduct high quality academic research. Also, thank you for making our research projects a priority, for indulging our discussions that tended to run beyond our meeting end time, and for being so supportive during my maternity leave. You've always given my research your undivided and meticulous attention, and I am profoundly grateful.

Second, thank you to the rest of my committee, Hong and Feng, for your guidance and support throughout this process. Hong, my thorough enjoyment of your Foundations of Strategy course, and the nuanced and interesting pieces of wisdom you provided us with each week in that course, motivated my desire to receive feedback from you more regularly on my research. Thank you for taking the time to meet with me, for providing me with invaluable feedback on my work, and, importantly, for your constant encouragement that I would succeed in my PhD journey. Feng, thank you for sharing your academic wisdom with me, for always finding time in your busy schedule to discuss my work, and for your positive and encouraging attitude.

Third, I sincerely thank my co-authors and HBS colleagues that have contributed to my academic experience. Thank you, Jodi, for working with me during the first few years of my program and to help improve my theory-writing, and thank you Maria, for being such a great role model and for picking up the baton with our project to see our research through to publication.

Thank you to Xiang from RCS, for always being there to help me through my statistical questions, and to Melissa, for stepping in on my projects while I was on maternity leave. Also, thank you to the TOM doctoral students and to the doctoral office (especially Jen and Marais) for their constant support.

Fourth, I am so grateful to my friends who were always there for me and for being a source of strength: Ryan, Rose, Moonsoo, Ehsan, Cady, Sarah, Paul, Erica, and Ron. Special thanks to Ariella (who I knew I wanted to be best friends with as soon as we sat together on a bus ride during the HBS program admit day) for introducing me to commitment devices (and for being an incredible friend!), to Kala, for letting me vent when I needed it, and to Emily and Nataliya, for always making our friendship a priority when I came to visit. Also, thank you to my book club ladies (Claire, Candace, Jen, Amanda, and Molly) for your encouragement, for reminding me that there's life beyond my program, and for being such bosses in general.

Finally, thank you to my family. I thank my husband, Alex, for being such a constant source of support, for reminding me that I was capable enough to complete this program when my confidence wavered, and for always picking up the slack when I needed to work nights and/or weekends. Thank you to my parents and family - Karen, Paul, Matthew, and KC- for always being supportive and encouraging of my passions, and thank you to my family in-law - Geert, Maria, Natasha, Tatiana, and Sabrina – for your support and for providing me with a quiet place to work. Finally, thank you to my daughter, Emmeline, for being my greatest source of joy and for teaching me that it is possible to work well on little sleep. I hope I can always make you proud.

Sourcing Monitors to Monitor Your Sourcing:
Strategies to Manage Shilling and Shirking by Supply Chain Auditors

Ashley Palmarozzo

Jodi L. Short

Michael W. Toffel

INTRODUCTION

Multinational companies (MNCs) have come under intense pressure from stakeholders such as activists, consumers, and shareholders to ensure adequate working conditions in their global supply chains, and MNCs can suffer significant reputational damage when media exposés reveal that their suppliers' poor practices—from blocking emergency exit doors to employing child slaves—harm workers. To manage these risks, MNCs engage in extensive auditing of suppliers. Such audits have a dual purpose: to satisfy the legitimacy demands of external stakeholders and to overcome information asymmetries between MNCs and their suppliers that create reputational risks for the MNCs' brands. While these dual purposes of supplier auditing are closely related, they stand in some tension with one another. Auditing used as an accountability mechanism to promote external perceptions of legitimacy might not overcome information asymmetries between MNCs and suppliers if it is deployed ceremonially to deflect criticism, rather than rigorously to yield thorough and reliable information (Power, 1997).

Auditing theory and practice long have been concerned with incentivizing high-quality audits, which we define as audits that yield comprehensive and accurate information. Most research on auditing has looked to independence as the lodestar of auditor performance. More independence is supposed to produce higher-quality auditing because it is assumed that auditors whose incentives are overly aligned with the entities they audit might not comprehensively and accurately assess the latter's conduct. This paper enriches that account by exploring the tradeoffs independence can entail, including agency problems as described below.

Auditor independence is typically proxied by the degree of formal separation between the auditor and the audited entity, mapped onto a sliding scale from first- to third-party auditors. First-party auditors are employed by the audited entity—in other words, they conduct internal self-audits of their own firms. They have the least independence because there is no formal separation between them and the entity they audit. Second-party auditors are typically employed by a business partner of the audited entity. This occurs when, for instance, the employees of a buyer firm or a franchisor audit the products or practices of a supplier or a franchisee. Second-party auditors have formal independence from the entities they audit because they are not employed by those entities, but they are employed by firms whose incentives might be aligned with those of the audited entities. Third-party auditors are employed by firms whose primary business is to conduct audits (audit firms); they have no formal employment relationship with the entities they audit or with the business partners monitoring those entities.

The emphasis on independence to incentivize high-quality auditing assumes that the core threat to auditor performance is shilling—covering up an audited firm's wrongdoing to mislead others into doing business with it. Financial auditors, for example, might mislead investors by covering up material misstatements on a company's financial statements or environmental auditors might mislead sustainability-driven consumers by covering up a company's polluting activities. Indeed, auditors do face incentives to shill for the firms they audit. First-party auditors are beholden for their livelihood to the firms they audit; their employers have leverage to pressure them into leniency (Coffee, 2004). Second-party auditors might be subject to similar pressures if the firm that employs them gains a competitive advantage from the audited partner's misconduct (Lebaron and Lister, 2015; McAllister, 2012)—for instance, if buyers benefit (at least in the short term) when their suppliers manufacture products more quickly and cheaply by evading occupational health and safety standards. Arguably, third-party auditors are not subject to these pressures since they are not employed by either the audited entity or the firm on whose behalf the audit is being conducted (in our context, the MNC buying from the audited supplier). Because third-party auditors have more formal independence than first- or second-party auditors, they are typically considered "the gold standard" (Prakash and Potoski, 2007, p. 790) in auditing. External stakeholders making legitimacy demands on MNCs do not trust MNCs to rigorously police their own suppliers and thus typically prefer third-party to second-party auditors (e.g., Lebaron and Lister, 2015; McAllister, 2012).

This basic story about how independence promotes high-quality auditing has serious limitations. First, it is not clear that the assumptions about the incentives and motives of the various actors are correct. For instance, although third-party auditors are not employed by the firms they audit (or the business partner beneficiaries of these firms), their relationships with their clients can create conflicts of interest that undermine performance. Research has shown that third-party auditors are more incentivized to overlook poor performance when they are directly compensated by the audited entity (Duflo, Greenstone, Pande, and Ryan, 2013; Jiang, Stanford, and Xie, 2012; Pierce and Toffel, 2013), when they have opportunities to cross-sell to it (Causholli, Chambers, and Payne, 2014; Pierce and Toffel, 2013), or when they audit it multiple times (Griffin and Tang, 2011; Lennox, Wu, and Zhang, 2014; Short, Toffel, and Hugill, 2016). Nor should it be assumed that firms monitoring business partners want auditors to shill for them. Indeed, such firms are often highly motivated to learn the truth about their partners' misconduct so that they can narrow information asymmetries in order to manage risks.

Second, the concept of independence does not indicate how to structure monitoring in multi-layered assurance relationships, such as those across global value chains and other contexts in which companies monitor business partners. If more than two entities are involved, it is not clear who should be independent from whom to promote high-quality auditing. For instance, a third-party auditor hired by an MNC to monitor its suppliers is formally independent of those suppliers. But so are the MNC's own employees, who could conduct second-party audits. Independence, on its own, does not tell MNCs whether monitoring will be performed better by second-party auditors or by third-party auditors.

Third, even if the traditional independence story captures important aspects of auditor incentives, that is only part of a much broader incentive landscape. For instance, auditors might also have incentives to shirk—that is, cut corners—to reduce costs. Like shilling, shirking can cause auditors to overlook critical information and produce incomplete and thus misleading audit reports. But independence does not mitigate the incentives that lead to shirking. Indeed, greater separation between the auditor and the audited entity might actually increase incentives for shirking, because the more independence auditors have, the more easily they can get away with it. In other words, while independence generates incentives that can improve auditor performance, it also creates agency problems that could undermine auditor performance.

The auditing literature lacks a theoretical framework for assessing the tradeoffs between independence and agency problems in incentivizing high-quality compliance auditing—by which we mean auditing that yields a comprehensive and reliable assessment of the audited entity's adherence to standards. This paper therefore proposes a novel conceptualization of auditor performance in the context of business partner auditing as a function not only of independence, but also of agency problems associated with sourcing decisions. Independence is created by outsourcing monitoring functions. While independence might create incentives that can enhance auditor performance, outsourcing creates agency problems that can reduce it. We synthesize insights from the literatures on auditor independence, agency theory, outsourcing decisions, and concurrent sourcing to explore this tradeoff between independence and agency problems and to theorize how it can be mitigated by other sourcing arrangements, including (a) insourcing, in which an organization uses its own employees to conduct audits of business partners; (b) concurrent sourcing, in which an organization outsources some audits of business partners in a given market and insources others; and (c) rotational sourcing, in which an organization rotates audit providers to ensure that subsequent audits of a given business partner are conducted by different auditing firms (audit firm rotation).

We explore these questions in the context of auditing global supply chain factories for compliance with labor standards contained in MNCs' supplier codes of conduct. In response to pressure from stakeholders, thousands of MNCs, including all *Fortune* 500 companies, have adopted codes of conduct that require suppliers to meet specified workplace standards (McBarnet, 2007). Many MNCs conduct audits to monitor and assess suppliers' adherence. This empirical context is particularly useful for exploring the issues discussed above because it entails complex relationships among buyers, suppliers, and auditors that require difficult choices about how firms monitor business partners to manage business, legal, and reputational risks.

We develop and test our hypotheses using a proprietary auditing database from a global branded apparel MNC that monitors suppliers' compliance with its labor code of conduct by both using its own employees (second-party auditors) and outsourcing some monitoring functions to third-party auditors. First, we offer competing predictions about whether high-quality auditor performance is likely to be better incentivized through greater independence or through greater control. We find that second-party auditors outperform third-party auditors, suggesting the limitations of auditor independence as the touchstone of performance. Second, we hypothesize and demonstrate that concurrent sourcing incentivizes third-party auditors to improve performance. Third, we examine rotation in sourcing arrangements and hypothesize that auditor performance is better when the auditor's firm did not perform the previous audit. We find that this rotational sourcing strategy mitigates the waning stringency of both third-party and second-party auditors and that this occurs above and beyond the known benefits of rotating individual auditors.

These results have important theoretical and managerial implications. In terms of theory, we extend auditor independence research by treating auditor performance quality as an issue not only of incentives relating to independence but also of agency problems. The literature on agency problems in auditing has primarily focused on those between audited firms and the stakeholders demanding the audits (e.g., boards, nongovernmental organizations), leading to calls for greater auditor independence between them. We theorize that under certain conditions, particularly when principals genuinely wish to close information asymmetries with their agents, independence can impede auditing quality.

Our study generates several managerial insights that can help firms strategically monitor business partners. First, whereas industry wisdom holds that third-party auditors are key to high-quality auditor performance (Lebaron and Lister, 2015), we show that, at least in our sample, MNCs' own employees—serving as second-party auditors—perform better on average. Second, we demonstrate that firms need not insource all supplier audits; rather, they can improve third-party auditing performance through a concurrent sourcing strategy in which second-party auditors conduct some of the audits in the same market. Third, MNCs can better monitor business partners by rotating the firms doing the monitoring (including their own).

LITERATURE REVIEW

MNC auditing of supplier adherence to labor codes of conduct is a form of nonfinancial auditing that occurs within an elaborate ecosystem of private transnational governance initiatives aimed at improving working conditions in global value chains. These initiatives grew up in response to perceived limitations of national governments in regulating MNCs in a globalized economy. In this vacuum, NGOs, unions, and labor activists developed strategies of "regulating firms across their supply chains through 'voluntary' standards (sometimes developed in cooperation with NGOs and unions), internal and external monitoring systems, and new sanctions and incentives" (O'Rourke, 2003: 2).

The standards on which transnational labor governance systems are based emanate from institutions including intergovernmental organizations such as the International Labor Organization, NGOs such as the Fair Labor Association, and standard-setting bodies such as ISO and Social Accountability International. The codes of labor conduct that MNCs require suppliers to adopt typically are based on standards from one or more of these sources. Many of these standard-setters also provide monitoring services, but such services are also offered by private, for-profit social auditing firms and nonprofit NGOs. Some MNCs employ their own in-house social auditors to monitor suppliers. MNCs impose codes of labor conduct to avoid reputational risk posed to their brands by poor supplier labor practices—for which they are not typically legally liable—and suppliers agree to them as a condition of doing business. Most MNCs that impose such codes use social auditors to monitor adherence.

This ecosystem of private transnational labor governance contains many types of auditors that could be deployed to monitor suppliers, but the literature provides limited tools for theorizing the performance implications of these choices—specifically, for predicting how insourcing, outsourcing, concurrently sourcing, and rotationally sourcing auditing functions affect auditor performance. Such questions do not present themselves in the financial auditing context, where companies have little choice of auditor type because the law requires third-party auditors to certify the accuracy of financial statements.

There is a substantial literature on nonfinancial auditing that addresses empirical contexts analogous to ours, including auditing for certification to management systems standards such as ISO standards for quality (9001), environment (14001), and occupational health and safety (45001); Hazard Analysis Critical Control Point (HACCP) standards for food safety; and sustainability standards promulgated by NGOs like the Forest Stewardship Counsel and Fair Trade. Although studies in this literature recognize the complexity of relationships in global assurance regimes, they provide little purchase on the question of which players are best positioned to perform comprehensive and accurate audits. Much research in this domain focuses on different questions entirely, such as whether auditing can improve compliance with standards (Distelhorst, Locke, Pal, and Samel, 2015; Lebaron and Lister, 2015; Locke, Qin, and Brause, 2007; Short, Toffel, and Hugill, 2020); what is an optimal auditing style (e.g., compliance vs. continuous improvement or cooperative vs. punitive) (Locke, Amengual, and Mangla, 2009; Power and Terziovski, 2007); what information obtained by auditors winds up in sustainability disclosures (Thomson 2007); and how auditing constructs rational myths of accountability (Boiral, 2003; Boiral and Gendron, 2011; Lebaron and Lister, 2015). Studies on nonfinancial auditing performance are highly influenced by the financial auditing model (Boiral and Gendron, 2011), tending to assume that auditing will be conducted by third-parties or that the only alternative is self-regulation (internal, first-party auditing) (e.g., Power and Terziovski, 2007), which has obvious negative implications for auditor performance. As a result, studies of nonfinancial auditing performance have focused exclusively on factors that influence the quality of outsourced (third-party) auditors without considering the performance implications of different sourcing arrangements.

Consequently, what we know about the factors that promote high-quality auditor performance in both financial and nonfinancial auditing comes exclusively from studies of third-party auditors. Driving forces identified in these literatures include independence, information advantage, expertise, institutional and market pressures, and cognitive biases. Research has shown that auditors produce assessments biased in favor of their audited clients when they are financially dependent on those clients; for instance, when the auditor is paid by the audited client (Duflo et al., 2013; Jiang, Stanford, and Xie, 2012); cross-sells non-auditing services to the audited client (Causholli et al., 2014; Pierce and Toffel, 2013) or has a long-term relationship with it (Griffin and Tang, 2011; Short, Toffel, and Hugill, 2016). Conversely, studies have shown that auditors perform better when they have informational advantages such as familiarity with the local market (Fan et al., 2020) or forms of expertise such as extensive auditing experience or superior professional credentials (Short, Toffel, and Hugill, 2016). Reputational or legal pressures in auditors' institutional environment can promote high-quality auditing (DeAngelo, 1981; Francis and Wilson, 1988; Lytton and McAllister, 2014). Social expectations and social dynamics related to auditors' gender likewise can shape audit outcomes (Short, Toffel, and Hugill, 2016). The evidence on market pressure in the form of competition is more mixed: some studies find that more competition in an audit market promotes higher-quality auditing (e.g., Boone, Khurana, and Raman, 2012; Huang, Chang, and Chiou, 2016); others find that it promotes lower-quality auditing (e.g., Bennett, Pierce, Snyder, and Toffel, 2013; Newton, Wang, and Wilkins, 2013), perhaps by enabling firms to shop for the most lenient auditors. Finally, research shows that cognitive biases, including bounded awareness, can limit auditors' ability to identify violations (Krogh, Roos, and Slocum, 1994; Kumar and Charkrabarti, 2012) and erode their performance in repeated audits (Short, Toffel, and Hugill, 2016).

A handful of studies explicitly address the performance implications of auditor selection processes, but these studies focus on the choice *among* third-party auditors—not between insourcing and outsourcing auditing functions—and investigate how the selection process can be designed to maximize auditor independence (e.g., Dao, Raghunandan, and Rama, 2012; Duflo et al., 2013; Lennox et al., 2014; Mayhew and Pike, 2004).

Second-party auditors have been largely ignored in both the financial and nonfinancial auditing literatures. This is understandable in the financial auditing literature, where the only relevant comparison is between first-party (internal) auditors and third-party (outside) auditors and firms are legally required to use third parties. It is less clear why second-party auditors are absent from the nonfinancial auditing literature. Although scholars have suggested that key players in global sustainability governance are aware that second-party auditors could play an important role in assurance regimes (e.g., Meidinger, 2019), there are, to our knowledge, no rigorous studies of their efficacy or of the incentive landscape they face. This is a significant blind spot because it is not clear that the incentive structures facing second- and third-party auditors are identical or that the insights from research on third-party auditor performance can be applied wholesale to second-party auditors. Specifically, third-party auditors have agency problems that might amplify incentives to shirk, creating a layer of incentives that has not yet been considered.

HYPOTHESES

Our study develops a framework for assessing the tradeoffs between independence and agency costs in incentivizing high-quality compliance auditing. Our framework is based on a simple but largely unexplored intuition: auditor independence is achieved by outsourcing monitoring to third parties. We argue that the choice among different types of auditor should be informed not only by incentives related to auditor independence and other factors identified in the literature, but also by the incentives activated by decisions about how to source auditing functions. We draw on studies in the literatures on agency theory, outsourcing, and concurrent sourcing to hypothesize when auditor performance benefits from more independence, when it benefits from more control, and how firms can mitigate the agency problems entailed in outsourcing monitoring.

Better performance through outsourcing

Auditor independence has long been the touchstone of audit quality across a wide range of monitoring contexts (Ammenberg, Wik, and Hjelm, 2001; Levitt, 2000; McAllister, 2012). Formal independence from the audited party is thought to attenuate auditor bias and avoid shilling for audited entities by reducing a key conflict of interest that would obtain in a first-party audit (Kouakou, Boiral, and Gendron, 2013)—namely, that first-party auditors stand to gain from their employer's malfeasance and thus have incentives to overlook it (Kraakman, 1986). This conflict of interest is also theorized to compromise the quality of second-party auditing; for example, when firms employ in-house auditors to monitor suppliers (Lebaron and Lister, 2015).

Such auditors might overlook suppliers' malfeasance if they believe that it provides competitive advantages, such as lower-cost supplies, to their employer. By contrast, third-party auditors, gaining no direct benefit from suppliers' malfeasance, may have less incentive to overlook it (Coffee, 2004).

In addition to theorized adverse incentives that undermine first-party or second-party auditor performance, some have theorized that reputational concerns incentivize high-quality performance by third-party auditors. This theory holds that third parties who monitor multiple clients will not jeopardize their reputation in the market for the potential gains of shilling for one client (DeAngelo, 1981; Kraakman, 1986). Some studies show that concern for reputation can indeed reduce the likelihood that monitors succumb to client pressure to hide wrongdoing (Francis and Wilson, 1988; Ljungqvist, Marston, Starks, Wei, and Yan, 2007; Pierce and Toffel, 2013). This should apply to supply chain auditing because third-party firms, which audit on behalf of many brands, are likely to use their reputations for high-quality audits to attract and retain clients.

We therefore hypothesize:

Hypothesis 1a: Third-party auditors outperform second-party auditors.

Better performance through insourcing

Agency theory and the risk of reputation spillovers suggest the opposite relationship: that second-party auditors will outperform third-party auditors. Agency problems arise when one party (the principal) delegates work to another party (the agent) and information asymmetry prevents the principal from observing the agent conduct the work. Because the incentives of the principal and agent are seldom fully aligned, a moral hazard problem often arises whereby the agent is motivated to shirk (Jensen and Meckling, 1976; Klein, Crawford, and Alchian, 1978). One solution has been to introduce a three-tiered agency model in which the principal hires a supervisor to monitor the agent's behavior. Studies have modeled how government agencies (principals) can assure that regulated entities (agents) comply with regulations by licensing "supervisor" companies to inspect them (Tirole, 1986). For example, many state governments authorize private garages to conduct emissions testing in order to ensure that vehicle owners meet emissions standards. However, this approach introduces an additional layer of agency problems between the principal and the supervisor, which can lead to supervisors colluding with agents rather than faithfully exercising the charge of the principal (e.g., Khalil and Lawarrée, 2006; Kofman and Lawarrée, 1993; Schneider, 2012; Seifter, 2006).

Agency problems between principals and supervisors are likely to be particularly severe in two situations, both of which have led firms to insource to better calibrate incentives. First, agency problems are exacerbated when it is more difficult for principals to monitor performance—especially when the quality of the products (Parmigiani, 2007) or services (Jia, 2018) supplied is not readily apparent. For example, firms are more likely to assign sales and distribution functions to their own employees, rather than outsource, when sales practices or measurement challenges make performance monitoring more difficult (Anderson and Schmittlein, 1984; John and Weitz, 1988).

Second, agency problems can be particularly severe when poor performance by an agent (or, in our context, a supervisor) creates greater reputational risk for the principal than for the agent or the supervisor itself. Such asymmetrical risk of reputational spillovers exacerbates the misalignment of incentives between principals and agents (or supervisors) because it prevents agents (or supervisors) from fully internalizing the costs of their poor performance. Studies have found that firms tend to insource functions they would ordinarily outsource when poor-quality work exposes them to high risk of reputational harm that is not fully internalized by their agents, so that they can better control the work and align the incentives of those performing it (Mayer, 2006; Nickerson and Silverman, 2003). Nickerson and Silverman (2003), for example, find that trucking companies tend to use drivers they employ rather than third-party drivers on routes for which poor delivery performance seriously imperils the company's reputation or its broader operations and when these risks are not fully internalized by the third-party drivers.

The social auditing context is characterized by both monitoring difficulty and the asymmetrical risk to principals of reputational spillovers. Supply chain audits are conducted because buyers have limited visibility into their suppliers' operations and the audits are conducted at factories around the world, often thousands of miles away from the buyer's operations. Moreover, brands face significantly more risk than do third-party auditors of reputational spillovers from exposés revealing suppliers' poor labor practices. Studies have demonstrated that the press and activists are more likely to target high-profile brands, rather than lesser-known companies, for their relationship with abusive suppliers because this strategy garners more attention and is more likely to provoke a response from the industry (Bartley and Child, 2014; King and McDonnell, 2015). As providers of business-to-business services, third-party auditors are not household names and are therefore unlikely to be tagged by the press or activists, even if their shoddy work failed to uncover poor conditions that led to worker harms.[1] Moreover, it is difficult for branded principals to impose reputational discipline on social

[1] Examples of typical media reports describing poor working conditions in global supply chains that blame factory owners and brands but do not (or only scarcely) mention auditors include articles about worker suicides at Apple supplier Foxconn (Merchant, 2017), the fatal Tazreen factory fire (Manik and Yardley, 2012), and the fatal Rana Plaza factory collapse (Rushe, 2016). Similarly, NGOs focusing on global supply chain working conditions, such as the Worker Rights Consortium and the Fair Labor Association, tend to fault brands; while they occasionally express skepticism about auditors, they are far more likely to name brands than audit companies.

auditors because, while financial auditors are required to issue public restatements of their mistakes, the work of social auditors is almost always confidential and brands do not routinely reveal publicly the identity of their auditors. There is no systematic way for an MNC buyer (the principal) to learn which social auditors make more mistakes and have been dismissed by other buyers. Consequently, even if third-party auditors have some incentive to up their game when they know their branded clients risk reputational spillovers, they are unlikely to fully internalize this risk because they do not face commensurate reputational pressure.

We argue that under these circumstances, a principal's superior control of its own employees, acting as supervisors, can do more than independence to produce high-quality auditor performance. We therefore propose the following alternative hypothesis:

Hypothesis 1b: Second-party auditors outperform third-party auditors.

Better performance through concurrent sourcing

Recognizing that there might be performance benefits to controlling auditors through insourcing prompts questions about whether a concurrent sourcing strategy could raise third-party auditor performance without entirely sacrificing the benefits of outsourcing. Concurrent sourcing occurs when a buyer purchases the same goods or services from both external and internal suppliers—in our context, when the MNC buyer uses both second- and third-party auditors to assess factories within a given market (though each individual audit is conducted by a single auditing firm). Drawing on both types of auditor bolsters the buyer firm's ability to control auditors through three mechanisms without insourcing all auditing.

First, the buyer firm can better monitor outsourced production. Several studies argue that concurrent sourcing enables buyers to learn from in-house suppliers about production costs, technologies, and metrics, enhancing their ability to negotiate with and monitor outsourced suppliers (Bradach and Eccles, 1989; Heide, 2003; Jia, 2018; Puranam, Gulati, and Bhattacharya, 2013). Moreover, concurrent sourcing has been theorized to help firms evaluate the performance of third-party suppliers and compare it to that of internal suppliers. For instance, concurrent-sourcing firms can tap in-house production knowledge to better assess whether poor products from a third-party supplier are due “to genuine manufacturing problems or to surreptitious supplier cheating” (Heide, Kumar, and Wathne, 2014: 1166). Firms can benchmark performance between in-house and third-party suppliers to “ratchet” up performance by creating “a virtuous cycle of continuous improvement” (Puranam et al., 2013: 1151). Similarly, the franchising literature has shown that company-owned units’ performance can be used as a benchmark to motivate franchisee-owned units to improve (e.g., Bradach, 1997).

Second, in the supply chain auditing context, concurrent sourcing places in-house auditors on the ground in the same markets as third-party auditors, which facilitates the MNC’s ability to monitor third-party auditors—just as physical proximity reduces a franchisor’s monitoring costs over its franchisee (Norton, 1988; Rubin, 1978). In particular, concurrent sourcing makes it easier for in-house auditors to conduct spot-checks of third-party auditors’ work by visiting the same factories. Even if such spot-checking is done sparingly to minimize duplicative auditing costs, the market presence of second-party auditors signals to third-party auditors that the MNC can readily assess their work, which should deter shirking and encourage higher-quality performance.

Third, concurrent sourcing provides a plausible threat that the buyer can readily increase its use of insourcing if it grows dissatisfied with the quality of third-party auditors, just as dual sourcing creates competition meant to improve the performance of two external suppliers (e.g., Klotz and Chatterjee, 1995; Yang, Aydin, Babich, and Beil, 2012). Insourcing thus acts as a form of competition to third-party auditors, which can deter shirking and bolster performance (Dutta, Bergen, Heide, and John, 1995; Heide et al., 2014; Puranam et al., 2013).

Because introducing second-party auditors into a market through a concurrent sourcing strategy disincentivizes shirking by third-party auditors, we therefore hypothesize:

Hypothesis 2: Concurrent sourcing improves the performance of third-party auditors.

Better performance through rotational sourcing

Another consideration is whether firms should vary their sourcing of auditors over time to combat the potential erosion in the performance of both second-party and third-party auditors as they return to audit the same entities. Such variation is commonly referred to in auditing scholarship and policy as auditor rotation. A substantial body of research demonstrates that ongoing relationships between individual auditors and the firms they audit can create biases that undermine performance (Jin and Lee, 2018; Lennox, Wu, and Zhang, 2014; Short, Toffel, and Hugill, 2016), prompting some regulators to mandate periodic rotation of individual auditors (e.g., Securities and Exchange Commission, 2003). Less known, however, are the implications of the repeated use of the same audit firm. In an extensive review of the auditor rotation literature, Lennox (2014: 103) concludes that "we still do not have a clear idea as to whether mandatory audit firm rotation would make audit quality better or worse" and that, beyond the effects of rotating individual lead auditors, we still do not know "whether audit firms *also* should be periodically changed." Our study extends the logic underlying studies on the rotation of individual auditors to hypothesize the performance implications of rotating audit firms.

It has been widely observed that individual auditors are less stringent when auditing entities with which they have ongoing relationships. Individual auditors who build such relationships with audited suppliers risk losing the emotional distance needed to be an effective monitor, which may lead them to become more sympathetic to a supplier's position (Moore, Tetlock, Tanlu, and Bazerman, 2006; Short, Toffel, and Hugill, 2016) or more vulnerable to bribery (Khalil and Lawarrée, 2006; Montiel, Husted, and Christmann, 2012). Repeated interactions between individual auditors and the suppliers they audit can also exacerbate the problem of "bounded awareness"—the suite of cognitive mechanisms that lead individuals to focus only on certain categories of information (Chugh and Bazerman, 2007). Consistent with this, one study found that when individual social auditors returned to the same supplier, they recorded fewer violations than did auditors who were new to the factory (Short, Toffel, and Hugill, 2016).

In addition to these *individual* emotional and cognitive factors shaping auditor performance, auditors are also affected by the *organizational* environment of the company in which they are embedded—whether it is a dedicated audit firm employing third-party auditors or a firm employing its own second-party auditors to monitor its business partners. Specifically, a company's organizational structures, culture, and routines (to which we refer collectively as "organizational structures") are likely to shape its employees' cognitive heuristics via mechanisms such as training (Metzger, Dalton, and Hill, 1993), organizational norms (Trevino, Butterfield, and McCabe, 1998), codes of conduct (Cressey and Moore, 1983), and social networks (Brass, Butterfield, and Skaggs, 1998). Thus, an auditor's performance is likely to decline not only when he or she has already audited a given entity, but also when coworkers have audited it. For example, auditors relying on organizational structures conveyed by their firm might, in consecutive factory visits, focus on similar violation types, over-search areas of the factory previously covered by their colleagues (where issues are more likely to have been resolved in response to prior audit findings), and under-search other areas that might have revealed violations that were not stressed in firm training materials.

Varying the firms from which audits are sourced should address these issues. Different companies are likely to have different organizational environments, lessening the danger of cognitive limitations in repeated visits. We therefore hypothesize:

Hypothesis 3: Auditor performance is better when the auditor's firm did not perform the factory's prior audit.

DATA AND MEASURES

Empirical context and measures

We obtained (via a data-sharing agreement) factory audit data from one MNC buyer, which we refer to RetailBrand to anonymize their identity. RetailBrand is a large retailer of apparel, home goods, and beauty products with over 300 stores worldwide. RetailBrand monitors its supply chain factories, in dozens of countries, to assess compliance with its supplier code of conduct, which is closely aligned with the Ethical Trading Initiative's base code of labor practices. RetailBrand requires annual audits of all supplier factories and requires the same audit process and scope for all audits, regardless of the sourced product. Each audit is conducted by either second-party auditors (its employees) or auditors from one of several third-party audit firms. Auditors review the supplier's documents, tour the factory to observe conditions, and interview workers.[2] Factories are then assigned an overall "social audit rating" (Red-Critical, Red, Amber, or Green) based on the number and severity of the violations.

RetailBrand selects the auditors and pays for all audits and the vast majority of both second- and third-party auditors are local to the supplier factory's country or region.[3] Ninety-five percent of third-party audits are conducted by for-profit audit firms, with the rest conducted by nongovernmental organizations, intergovernmental organizations, or firms whose name in the dataset was blank. Third-party auditors also conduct audits for other MNC brands, whereas second-party auditors only conduct audits for RetailBrand.

Our unit of analysis is an audit of a factory conducted on a particular day (a factory audit). Because our model uses lags of several independent variables (information from the prior audit, which is missing for a factory's first audit), our estimation sample excludes each factory's first audit in our dataset. Our sample thus includes 11,099 audits for RetailBrand of 3,289 factories in over 40 countries during 2007–2017. Second-party auditors conducted 28 percent (3,078 audits) of these; third-party auditors conducted the remaining 72 percent (8,021 audits). RetailBrand does not own any of these factories and nearly all of them supplied other buyers when they were audited for RetailBrand.

Dependent variable

We measure the extent to which factories comply with the MNC code of conduct as the number of *new violations* recorded during an audit, similar to several studies that predict noncompliance with standards governing working conditions (Short, Toffel, and Hugill, 2016),

---

[2] The number of auditor-days an audit requires is determined by the factory's size (measured by number of workers excluding management), in accordance with Sedex Members Ethical Trade Audit (SMETA) methodology.
[3] RetailBrand selects third-party auditors based on their capabilities, capacity, geographic coverage, skills profile of auditors, and other factors.

food safety (Ibanez and Toffel, 2020), and air pollution (Bennett et al., 2013; Duflo et al., 2013; Pierce and Toffel, 2013; Schneider, 2012). But whereas those studies used total violation counts, including both newly recorded violations and any unresolved violations identified in prior audits, we use only the newly recorded violations in order to (a) isolate the focal auditor's ability to identify new issues and (b) avoid the potential confound of a factory's unobserved heterogeneous ability or willingness to resolve previously identified violations.

The code of conduct contains about 90 binary compliance elements that span over 10 violation categories that track the Ethical Trading Initiative's Base Code (itself based on International Labour Organization core conventions), and include categories such as: child labor, working conditions, and worker discrimination.

Independent variables

To measure the level of independence and control between the MNC buyer and the auditor, we coded *second-party auditor* as 1 when an audit was conducted by a second-party auditor (the MNC's employees) and as 0 when conducted by a third-party auditor.

We identified factory audits occurring in a concurrent sourcing market by coding *concurrent source* as 1 when second-party auditors conducted between 10 and 90 percent of all audits in a particular country-year, and 0 otherwise.[4] For example, if 250 of the 1,000 factory audits in China in 2015 (that is, 25 percent) were conducted by second-party auditors, we would

[4] Our definition of concurrent sourcing is akin to that in studies of companies that owned some restaurants and franchised others (Bradach 1997) and of trucking companies that assigned some routes to employee drivers and others to third-party drivers (He and Nickerson, 2006).

code *concurrent source* “1” for all 1,000 of them. This 10-percent-to-90-percent range is consistent with other empirical studies of concurrent sourcing (e.g., Parmigiani, 2007).[5]

We constructed three binary variables to measure the independence between the audited factory and the audit firm. We operationalize rotational sourcing by creating *different audit firm*, which indicates that a factory audit was conducted by a firm other than the one that conducted that factory’s prior audit. This includes scenarios in which one third-party audit firm replaces another, when a third-party audit firm replaces the buyer firm’s second-party auditors, or when the buyer firm’s second-party auditors replace a third-party audit firm. *Different lead auditor (same audit firm)* indicates an audit that was conducted by the same audit firm that conducted the factory’s prior audit but with a different lead auditor. *Same lead auditor (same audit firm)* indicates an audit that was conducted by the lead auditor and audit firm that conducted the factory’s prior audit.[6] Collectively, these variables enable us to measure the effect of audit firm rotation while accounting for lead auditor rotation, which prior research demonstrates can itself affect the number of violations cited (Short, Toffel, and Hugill, 2016).

Control variables

We created a number of audit-level variables to control for factors that might affect auditor performance. *Unannounced audit* equals 1 when the audited factory had no advanced notice of the audit, and 0 otherwise.[7] Factories receiving advanced notice have more time to

[5] That is, *concurrent source* equals 0 when either second-party auditors or third-party auditors conducted nearly all (more than 90 percent) of the audits in a country-year. Robustness tests that measured concurrent sourcing using several alternative ranges (1% to 99%, 5% to 95%, 15% to 85%, 20% to 80%, and 25% to 75%) produced nearly identical results, as discussed below and reported in Table A-5 in the Appendix.
[6] We identified 70 audits for which the audit firm rotated but the lead audit did not, which we assume to be coding errors. We constructed a binary variable to flag these instances and include it in all models.
[7] For the 260 audits for which *unannounced audit* was missing, we coded these missing values “0” and included in our models a separate dummy variable coded “1” to designate those codings and “0” otherwise. Re-estimating our models excluding these 260 audits yielded nearly identical results.

resolve or conceal potential new violations, which might reduce the number of violations recorded.

We created several binary variables to designate audit scope, which could affect the number of violations recorded. *New factory audit* refers to a new supplier factory audit, conducted before the MNC buyer places an order. *Limited scope audit* refers to an audit focused on a new addition to a factory or on a new supplier factory whose prior audit was incomplete.[8] *Re-audit* refers to a routine audit conducted roughly annually after the factory's new factory audit or prior re-audit. *Follow-up audit* refers to an audit typically conducted a few months after an audit has yielded a "Red" or "Red-Critical" overall social audit rating. We created these variables for both the factory's focal audit and its prior audit.[9]

Because research has found that auditor gender can influence performance (Short, Toffel, and Hugill, 2016), we created a continuous variable, *female lead auditor*, that reflects the probability that a lead auditor's first name is a female name in the audited factory's country.[10] We used genderize.io, which uses profile information across major social networks to predict gender and has been used in other scholarship for this purpose (e.g., Greenberg and Mollick, 2017; Lundberg and Stearns, 2019; Rubineau and Fernandez, 2015).

[8] For example, auditors may need to return to a new supplier factory if it does not have all documentation prepared on their first visit.
[9] Because our model uses some lagged independent variables that result in our estimation sample excluding each factory's first audit, no focal audits are *new factory audits*. For the 610 audits for which audit scope was missing, we coded all audit type variables "0" and created a separate dummy variable coded "1" for these 610 audits and "0" otherwise.
[10] For the 1,274 audits for which a gender could not be predicted based on the lead auditor's first name, we coded *female lead auditor* "0" and included in our models a separate dummy variable coded "1" to designate those codings and "0" otherwise. Our data includes the names only of lead auditors and not of other audit team members and does not include gender.

We also measured factory-level characteristics that might influence auditor performance. We created *percent factory supplied to MNC buyer* as the proportion of a factory's output supplied to the MNC buyer at the time of the focal audit, which might influence how beholden it is to that buyer and how motivated it is to adhere to the code of conduct.[11] Because factories with more employees might have more violations, we create *total workers (log)* as the natural log of the number of factory workers (plus 1) at the time of the audit.[12]

Table 1.1 reports summary statistics and Appendix Table A-1 reports correlations.

Table 1.1: Summary statistics

	Mean	SD	Min	Max
New violations	4.57	4.94	0	67
Second-party auditor	0.28	0.45	0	1
Concurrent source	0.91	0.28	0	1
Auditor rotation				
Different audit firm	0.56	0.50	0	1
Different lead auditor (same audit firm)	0.28	0.45	0	1
Same lead auditor (same audit firm)	0.16	0.37	0	1
Female lead auditor	0.39	0.46	0	1
Unannounced audit	0.23	0.42	0	1
Total workers	493	869	2	10,601
Total workers (log)	5.40	1.23	1.10	9.3
Audit type				
Limited scope audit	0.13	0.33	0	1
Re-audit	0.52	0.50	0	1
Follow-up audit	0.30	0.46	0	1
Percent factory supplied to MNC buyer	0.33	0.24	0.000003	1
Audit sequence	4.02	1.97	2	8
Audit year	2014	1.90	2007	2017
N (audits)		11,099		

[11] For the 3,197 audits for which *percent factory supplied to MNC buyer* was missing or zero, we coded these missing values as the variable's mean and included in our models a separate dummy variable coded "1" to designate those codings and "0" otherwise.
[12] For the 498 audits for which *total workers (log)* was missing or zero, we coded these missing values as the variable's mean and included in our models a separate dummy variable coded "1" to designate those codings and "0" otherwise.

EMPIRICAL ANALYSIS

Empirical model

We estimate the extent to which our hypothesized variables influence auditor performance by predicting factory noncompliance recorded in an audit as a function of our independent variables, controlling for the aforementioned factors that might affect the factory's on-the-ground working conditions. Our approach mirrors that of other studies of auditor and inspector performance (e.g., Bennett et al., 2013, Duflo et al., 2013; Ibanez and Toffel, 2020; Pierce and Toffel, 2013; Schneider, 2012; Short, Toffel, and Hugill, 2016). Specifically, we estimate the following model:

$$Y_{id} = F\left(\beta_1 X_{id} + \beta_2 \lambda_{id} + \beta_3 \rho_{i,d-1} + \beta_4 \alpha_i + \beta_5 \delta_y + \beta_6 \gamma_s + \epsilon_{id}\right) \quad (1)$$

Y_{id}, is *new violations* recorded in audit *d* of factory *i*. *F()* represents the Poisson function. X_{id} represents our hypothesized explanatory variables: *second-party auditor* (H1), *concurrent source* (H2), and *different audit firm* (H3). λ_{id} represents the control variables associated with the focal audit that were described above (*unannounced audit*, *same lead auditor (same audit firm), limited scope audit*, *follow-up audit, female lead auditor, percent factory supplied to MNC buyer,* and *total workers (log))*. $\rho_{i,d-1}$ refers to factors associated with the factory's prior audit that we control for: *new factory audit (prior audit), limited scope audit (prior audit)*, and *follow-up audit (prior audit). Different lead auditor (same audit firm)* serves as the omitted category for *different audit firm* and *same lead auditor (same audit firm)*. The focal (and prior) audit versions of *re-audit* serve as the omitted category for the focal (and prior) audit versions of *new factory audit, limited scope audit*, and *follow-up audit*.

α_i captures supplier-factory fixed effects, which control for the factory's industry, location, and other time-invariant factory-specific effects, such as a factory's general willingness or ability to comply with audits and the factory's institutional context. δ_y, refers to audit-year fixed effects, which control for secular changes that might affect factory violations or auditor stringency. We also control for audit sequence—whether the audit is the factory's first audit, second, and so on—to account for the possibility that being audited provides factories the opportunity to learn how to comply. We do so by creating a dummy variable for each audit sequence value (γ_s) after first winsorizing (top-coding) the sequence count at its 95th percentile (eighth audit) to mitigate the influence of outliers, following an approach used by others (e.g., Bird, Short, and Toffel, 2019).

Identification and results

We estimate the models that test our hypotheses using quasi-maximum likelihood estimation of the Poisson likelihood function, which is consistent even for dependent variables that are not Poisson-distributed (Cameron and Trivedi, 1998).[13] We report results in Table 1.2.

[13] This estimation methodology also avoids the incidental parameter problem exhibited by regression models suited for count data with many fixed effects, such as a negative binomial model with fixed effects.

Table 1.2: Poisson Regression Results

Dependent variable = *new violations*		(1)		(2)		(3)	
		Coef	AME	Coef	AME	Coef	AME
H1	Second-party auditor	0.454**	2.2			0.183**	0.9
		(0.066)				(0.032)	
H2	Concurrent source			0.188**	0.8	0.149+	0.6
				(0.073)		(0.083)	
H3	Different audit firm	0.332**	1.6	0.303**	1.3	0.254**	1.1
		(0.082)		(0.041)		(0.026)	
	Same lead auditor (same audit firm)	-0.509**	-2.5	-0.253**	-1.1	-0.363**	-1.5
		(0.123)		(0.074)		(0.044)	
	Female lead auditor	0.069	0.3	0.125**	0.6	0.160**	0.7
		(0.074)		(0.046)		(0.028)	
	Unannounced audit	0.204*	1.0	0.150**	0.7	0.113**	0.5
		(0.094)		(0.050)		(0.029)	
	Total workers (log)	0.135	0.7	0.067	0.3	0.067+	0.3
		(0.097)		(0.044)		(0.036)	
	Limited scope audit	-0.829**	-4.1	-0.623**	-2.8	-0.690**	-2.5
		(0.215)		(0.093)		(0.060)	
	Follow-up audit	-0.513**	-2.5	-0.462**	-2.0	-0.470**	-2.0
		(0.073)		(0.043)		(0.028)	
	Percent factory supplied to MNC buyer	0.124	0.6	0.036	0.2	0.023	0.1
		(0.225)		(0.074)		(0.073)	
	New factory audit (prior audit)	0.303	1.5	-0.074	-0.3	-0.140*	-0.6
		(0.225)		(0.052)		(0.064)	
	Limited scope audit (prior audit)	-0.035	-0.2	-0.060	-0.3	-0.066	-0.3
		(0.154)		(0.042)		(0.046)	
	Follow-up audit (prior audit)	-0.065	-0.3	-0.125**	-0.6	-0.077**	-0.3
		(0.080)		(0.047)		(0.027)	
	Factory fixed effects	YES		YES		YES	
	Audit-year fixed effects	YES		YES		YES	
	Audit-sequence fixed effects	YES		YES		YES	
	Sample	Matched sample		Third-party audits		All Audits	
	N (focal audits)	1,241		7,009		10,104	
	Number of supplier factories	490		1,847		2,321	
	Sample average of *new violations*	5.023		4.295		4.567	

Table 1.2: Poisson Regression Results (continued)

Notes: Poisson coefficients with robust standard errors clustered by supplier factory for Models 1 and 3 and by supplier factory's country for Model 2. + p<0.10, * p<0.05, **p<0.01. AME is average marginal effects. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm).* Omitted category for *new factory audit, limited scope audit*, and *follow-up audit* is *re-audit*.

Comparing second- and third-party auditors' performance (H1a/b)

For Equation 1 to yield an unbiased estimate of the effect of an audit being conducted by a *second-party auditor* (versus a third-party auditor) on *new violations* (Y_{id}), no unobserved variables should be correlated with both Y_{id} and the decision to assign a second- versus third-party auditor. We questioned whether this assumption held in our data for a few reasons. First, our interviews with RetailBrand employees revealed that it prioritizes assigning second-party auditors to new factory audits. Second, exploratory analysis revealed that (a) second-party auditors were more likely to be assigned to audit factories that had worse prior audit results[14] and that (b) factories' prior and focal audit results were correlated. Second-party auditors may therefore be more likely to be assigned to factories with more violations, which confounds our ability to compare the performance of second- and third-party auditors.

To address this concern, we developed a matched sample of audit pairs—one audit assigned to a second-party auditor, the other to a third-party auditor—that were otherwise as similar as possible in terms of their prior audit result and of factors that might be correlated with both auditor assignment and our outcome variable.[15] Specifically, we used coarsened exact

[14] This finding is consistent with the transaction cost economics literature's prediction that firms internalize transactions that can impose larger externalities on the rest of the firm's operations in the form of reputational spillovers (Mayer, 2006; Nickerson and Silverman, 2003).

[15] We cannot instead control for the lagged number of new violations due to a dynamic panel problem and to our relatively short panel, which prevents us from using dynamic panel estimators. Only 34% of suppliers in our dataset had at least five audits, which is the minimum panel length to enable the second and third lagged dependent variables to serve as instruments.

matching[16] to exactly match on (a) a coarsened version of *new violations* recorded in the factory's prior audit, (b) the factory's prior *social audit rating*, (c) *unannounced audit*, (d) audit scope (*limited scope audit*, *re-audit*, or *follow-up audit*), (e) *concurrent source,* (f) the factory's country, (g) *audit year*, and (h) *audit sequence* (top-coded at the 95th percentile, the eighth audit).[17] The resulting matched sample includes 2,056 audits of 1,299 factories.[18] We assessed balance across 24 variables: 6 predictors of assignment to address endogeneity concerns and 18 predictors of performance to ensure common support across second- and third-party audits.[19] Comparing group means between second- and third-party audits in this matched sample indicates that 20 of the 24 variables are balanced (meaning that the group means were statistically indistinguishable at the 10-percent level), a substantial improvement over the mere one that was balanced in the full sample.[20, 21] We controlled for these unbalanced variables in our regression model and in a robustness model.[22]

[16] To implement our matching procedure, we rely on the *cem* command in Stata, which implements the method articulated in Iacus, King, and Porro (2008).
[17] We include the supplier factory's prior audit's (1) social audit rating and (2) number of new violations. While an audit's social audit rating correlated with the number of violations recorded in the audit, the social audit rating also accounts for violation severity.
[18] For summary statistics of this matched sample, see Table A-2.
[19] For completeness, Table A-3 compares balance along four additional variables: *total workers* (measured in number of workers, whereas we use the log value in our models) and three dummy variables denoting instances in which *unannounced audit, audit scope (focal audit)*, or *audit scope (prior audit)* was missing.
[20] Table A-3 reports the covariate balance of the matched sample and compares it to that of the full sample. p-values were obtained from t-tests for non-binary variables and from proportional tests for binary variables. The average standardized bias, which measures balance in terms of the similarity in covariate distributions, is 0.46 for the matched sample, a substantial improvement over 9.50 for the unmatched sample. Standardized bias is calculated as the difference in group means divided by the average standard deviation. Values closer to zero imply greater similarity in distribution.
[21] The kernel density graphs in Figure A-9 in the Appendix show that in the matched sample, the distribution of the variables for second-party audits and for third-party audits is very similar.
[22] Our primary models control for the three of these four variables that remained unbalanced: *different audit firm, different lead auditor (same audit firm),* and *same lead auditor (same audit firm)*. A robustness test model that includes the fourth unbalanced variable, *second-party auditor (prior audit),* yields nearly identical results (not reported).

We used this matched sample to test H1 and report the results as Model 1 of Table 1.2, clustering standard errors by factory.[23] The positive coefficient for *second-party auditor* (β=0.45; $p<0.01$; IRR=1.57) reveals that second-party auditors, on average, record 57 percent more new violations than third-party auditors; this supports Hypothesis 1b, which predicted that second-party auditors outperform third-party auditors. The average marginal effect of 2.2 indicates that while third-party audits in the H1 sample record an average of 4.3 new violations, second-party auditors record an average of 6.5.

We conducted several robustness tests to explore the sensitivity of our main H1 results. First, we assessed the extent to which our results were driven by the matched data sample on which they were estimated by re-estimating Model 1 on the full sample of audits. Second, to assess whether our results are sensitive to how performance is measured, we re-estimated Model 1 using three alternative dependent variables: *new violations top-coded at the 95th percentile* (19 new violations) to compress outliers, *major new violations* to assess whether results hold even after omitting minor violations (using RetailBrand's definitions of major and minor), and *total violations,* the metric more commonly used in the compliance literature. These four models yield coefficients on our hypothesized variables that are very similar in magnitude and direction to our primary results,[24] indicating that our H1 results are robust to this alternative sample and to these alternative metrics. Third, because Model 1 uses factory fixed effects to control for time-invariant factors, the model is identified only on factories with at least two audits. Only 1,241 (60%) of the 2,056 audits in the matched sample are associated with factories with at least two audits and thus are identified in our factory fixed-effects model. To ensure that this sample

---

[23] The *concurrent sourcing* variable drops out here because in this matched sample, the variable is (unsurprisingly) always coded 1 and thus exhibits no variation.
[24] Results are reported in Appendix Table A-4.

diminution is not driving our results, we re-estimated our model using two alternatives to our fixed-effects approach that do not drop any observations: one omits the fixed effects (i.e., a pooled model) and the other uses random effects instead of fixed effects. (In both cases we continue to cluster standard errors by factory.) Both approaches yielded coefficients statistically significant at the one-percent level and similar in magnitude to those of our main approach.

Assessing how third-party auditors' performance differs with concurrent sourcing (H2)

When testing H2, we restrict our sample to audits conducted by third-party auditors for which we have data for all variables in Equation 1.[25] To isolate the effect of transitioning from a third-party market to a concurrent sourcing market, we exclude the few instances (75 audits in 7 country-years) in which the reverse transition occurred—a supplier factory transitioned from a concurrent sourcing market to a third-party market. This results in an estimation sample of 7,967 third-party audits. Because Equation 1 includes factory-level fixed effects, the coefficient on *concurrent source* pertains to the average change—*within* a given factory—in new violations reported by third-party auditors between those years when the factory was *not* in a concurrent source market (that is, a year in which the MNC buyer only—or predominantly—relied on third-party auditors in that country) to when it *was* in a concurrent-source market (in which second-party auditors conducted between 10 and 90 percent of audits in the country-year).[26]

We hypothesize and find that third-party auditors record more *new violations* within the same factories over time once those factories are in a concurrent sourcing market. The principal

[25] Summary statistics of this sample are reported in Appendix Table A-2.
[26] Our identifying assumption is that this market transition was not sparked by unobserved factors that might have simultaneously affected both the decision to begin concurrently sourcing in a given market (country-year) and an erosion in factory working conditions in that market. Interviews with the MNC reveal that its decision to have its second-party auditors enter a particular country or to scale up its auditing in particular countries was instead based on growing order volume procured from the suppliers in that country.

endogeneity threat here is therefore that an unobserved factor might simultaneously drive both (a) worsening underlying working conditions in the factories in a given country and (b) the MNC buyer's propensity to begin concurrently sourcing audits in that country (that is, to begin assigning second-party auditors to some factories in that country). We are aware of no evidence to suggest that such an unobserved factor is present, as the MNC buyer tended to enter markets to begin concurrently sourcing once its purchasing volume from factories in that country merited the effort and investment to hire second-party auditors—a factor unrelated to anticipated improvements in overall working conditions in factories in that market.

Model 2 of Table 1.2 reports the results of testing H2, clustering standard errors by country because *concurrent source* is measured at the country-year level. The positive coefficient on *concurrent source* ($\beta=0.19$; $p<0.01$; IRR=1.21) indicates that third-party auditors record an average of 21 percent more new violations per audit after the MNC has deployed concurrent sourcing in the factory's market, which supports H2. The average marginal effect indicates that third-party audits in concurrent source markets average 3.7 new violations—0.8 more than the 2.9 average in markets served exclusively or nearly exclusively by third-party auditors.

We found similar results when, as robustness tests, we identified concurrent sourcing markets using five alternative thresholds (1 to 99 percent, 5 to 95 percent, 15 to 85 percent, 20 to 80 percent, and 25 to 75 percent) instead of our primary 10-to-90-percent approach.[27] Moreover, we find similar results when we re-estimate Model 2 using the three alternative dependent variables listed above (*new violations top-coded at 95th percentile*, *major new violations*, and *total violations*).[28] Collectively, these results indicate that our primary results are robust to

[27] These results are reported in Appendix Table A-5.
[28] These results are reported in Appendix Table A-6.

alternative thresholds when measuring our independent variable and to alternative measurements of the dependent variable.

Assessing how auditors' performance differs with audit firm rotation

To obtain an unbiased estimate of the effect of audit firm rotation on auditor performance (H3), the decision to assign a different auditing firm to a supplier factory's focal audit should be uncorrelated with the factory's prior audit results. Interviews with the MNC reveal that the assignment of audit firms to factory audits (and thus rotation) is based on auditor availability, which is unlikely to be correlated with prior audit results. Thus, we test H3 on the entire sample of audits for which we had data on all measures described above.

The results of Model 3, which tests H3, are reported in Column 3 of Table 1.2, where we cluster standard errors by factory. The coefficient on *different audit firm* ($\beta=0.18$; $p<0.01$; IRR=1.20) indicates that audits with rotated audit firms report 20 percent more new violations than those with rotated lead auditors within the same firm (the omitted category). The average marginal effect indicates that audits conducted by firms other than that which conducted the factory's prior audit report an average of 0.9 more new violations than the 4.0 average for audits conducted by the same firm as last time but with a different lead auditor.

The principal endogeneity threat in testing H3 is that an unobserved factor might simultaneously (a) worsen the working conditions of the factory, generating more new violations, and (b) prompt the MNC buyer to assign that factory to a different audit firm. We are not aware of plausible examples of such an unobserved factor, but as a robustness test, we re-estimated Model 3 on two alternative subsamples in which such endogenous audit-firm assignment is especially unlikely. First, we omitted from the sample those audits for which intentional audit firm assignment might have been especially tempting due to an unusually steep decline in *new violations*.[29] Second, we used coarsened exact matching to create a matched sample that matched one audit conducted by an audit firm that had not conducted the factory's prior audit (audit firm rotation) to an audit conducted by the audit firm that had conducted it (not rotation).[30] Re-estimating Model 3 on these two subsamples continues to yield positive coefficients on *different audit firm* of similar magnitude that are statistically significant at the one-percent level,[31] suggesting that our primary estimates of H3 are robust to these two scenarios of possible endogenous audit firm rotation.

Re-estimating Model 3 to instead predict each of the alternative dependent variables described earlier (*new violations top-coded at 95th percentile, major new violations,* and *total violations*) indicates that the relationship predicted by H3 is robust across these alternative metrics.[32]

As an extension, we explore the extent to which audit firm familiarity might attenuate the benefits of audit firm rotation, compared to rotating to an audit firm that has never audited the factory before. To do so, we decomposed *different audit firm* into *new different audit firm*—which equals 1 for an audit by a firm that has never audited the factory before and 0 otherwise—

[29] Specifically, we created an improvement score by calculating the percent reduction in the number of new violations, comparing a factory's prior audit to its preceding two audits (or single preceding audit when two were not available). Of the 5,099 audits with non-missing improvement scores, we omitted the 1,278 whose percent declines were the highest 25th percentile (improvement of 80% or greater). We supplemented the remaining 3,821 audits with the 6,000 audits for which we could not calculate improvement—when the focal audit was the factory's second audit or when both of the previous two audits yielded 0 new violations—yielding a subsample of 9,821 audits of 3,289 factories.
[30] Specifically, we matched on a coarsened version of new violations (prior audit), social audit rating (prior audit), second-party auditor (vs. third-party auditor), audit scope (new factory audit, limited scope audit, re-audit, and follow-up audit), supplier country, audit sequence (top-coded at 95th percentile), and audit year. The resulting matched sample contains 3,842 audits of 1,829 factories.
[31] These results are reported in Columns 1–2 of Appendix Table A-7, in which the coefficients on *different audit firm* are 0.24 and 0.30, somewhat larger than the 0.18 estimate from our main approach.
[32] These results are reported in Columns 3–5 of Appendix Table A-7.

and *familiar different audit firm*—which equals 1 when the audit firm has conducted audits at the factory before but not the most recent audit. We estimate this model on all audits and cluster standard errors by factory. The results indicate that, compared to audits conducted by the same firm as last time but with a different lead auditor (the omitted category), audits rotated to a *new different audit firm* report an average of 1.3 more new violations and audits rotated to a *familiar different audit firm* report an average of 0.7 more.[33] A Wald test comparing the coefficients on these variables confirms that their difference is statistically significant at the one-percent level (χ^2=14.6; p<0.01).

Control variables

Finally, we reflect on results associated with several control variables in Table 1.2. *Unannounced audits* yield an average of 0.5 to 1.0 more new violations than pre-announced audits, which is directionally in line with prior studies (Lebaron and Lister, 2015; Short, Toffel, and Hugill, 2016). Audits with a *female lead auditor* tended to report an average of 0.3 to 0.7 more new violations than those with male lead auditors, also consistent with the literature (Short, Toffel, and Hugill, 2016). Moreover, audits conducted by the same lead auditor who conducted the factory's prior audit report an average of 1.1 to 2.5 fewer new violations than audits led by a different lead auditor from the same audit firm, consistent with studies extolling the benefits of rotating lead auditors of the same audit firm (Lennox, 2014; Lennox, Wu, and Zhan, 2014; Short, Toffel, and Hugill, 2016). As expected due to their more circumscribed scope, *follow-up audits* and *limited scope audits* yield fewer new violations. We find no evidence that the number of new violations is affected by *total workers (log)* covered by the audit, recalling that our inclusion

[33] These results are reported in Appendix Table A-8.

of factory fixed effects means that this variable identifies only within-factory variation during our sample's few years. We also find no evidence that the number of new violations is affected by variation in within-factory *percent factory supplied to MNC buyer*.

DISCUSSION

Our study of auditing in the context of multi-layered supply chain assurance regimes reveals that auditor performance is shaped not only by the potential biases and conflicts of interest traditionally addressed through independence criteria, but also by the incentives to behave opportunistically common to all suppliers of goods and services. Our approach allows us to identify novel audit sourcing arrangements—rotating audit firms and controlling auditors through insourcing or concurrent sourcing—that can improve auditor performance. Our work makes five significant contributions to the literatures on auditor independence and outsourcing decisions and has important managerial implications, all discussed below.

First, by considering a range of sourcing options—insourcing, outsourcing, concurrent sourcing, and rotational sourcing—we identify cross-cutting issues that can affect auditor performance and we generate new insights about how to incentivize better auditor performance. While the focal problem in the auditor independence literature is auditors *shilling* for their clients, agency theory suggests that firms should also worry about third-party auditor *shirking*. Independence is not the solution to auditor shirking and, indeed, might exacerbate it. We show that, in our context, a firm can elicit better performance from auditors by controlling them rather than keeping them at a distance.

To be clear, we do not claim that second-party auditors always perform better than third-party auditors. Second-party auditors are likely to perform well only if their employer is committed to learning the truth about misconduct by its business partners. If their employer is committed to covering up such misconduct, control is likely to facilitate that. Thus, the most important scope condition on our findings is the principal firm's motivation to close information asymmetries. We note, however, that our finding that some firms exercise control of their auditors to generate more rather than less information is, in itself, an important contribution that should open new conversations in a field in which stakeholders tend to mistrust the motives of MNC buyers and thus reflexively demand third-party audits (e.g., Lebaron and Lister, 2015). Beyond motivation, as we discuss above, we also suspect that control is likely to be particularly beneficial for the many firms, such as branded MNC buyers, that face high risk of reputational spillover from suppliers' mistakes. However, data limitations prevent us from fully exploring this boundary condition. Research exploiting data from multiple MNC buyers is needed to identify and test these and other circumstances that might differentially affect the performance of second- versus third-party auditors.

Second, our study provides a novel window into the role of second-party auditors in assurance regimes. Second-party auditors have been an afterthought in a literature focused almost exclusively on first- and third-party auditors. This is likely because financial auditing provides the model for designing audit regimes (Power, 1997) and, in that context, firms are legally required to outsource audits to third parties rather than perform first-party audits of their own financials. But even scholars and advocates focused on more complex assurance regimes involving multi-layered business relationships tend to assume that second-party auditors suffer from the same biases and conflicts of interest as first-party auditors (e.g., Lebaron and Lister, 2015; McAllister, 2012). To our knowledge, our study is the first to empirically test this assumption and we find it to be mistaken—or at least incomplete. Our findings suggest that knowledge established in studies of third-party auditors should not be applied automatically to second-party auditors. Rather, it is necessary to theorize and test how incentives might shape second- and third-party auditor performance differently in different settings.

Third, our findings on concurrent sourcing and rotational sourcing suggest that auditor sourcing is not a binary choice between insourcing and outsourcing or between independence and control. Rather, sourcing strategies can be combined or allocated in different ways with implications for auditor performance. Understanding that auditor performance is partially a function of sourcing decisions reveals innovative opportunities to improve it. Of particular note, our finding that the competition generated by concurrent sourcing raises auditor performance is in some tension with studies in the auditing literature finding that competition can depress auditor performance (e.g., Bennett et al., 2013; Newton, Wang, and Wilkins, 2013). We suspect that this discrepancy can be explained by the opportunities for auditor shopping in a given context and the motivations of the firms that select the auditors. For instance, Bennett et al. (2013) theorize that competition in auto emissions testing dampens monitoring quality because it offers motorists who wish to conceal their vehicles' noncompliance more opportunities to shop for lax monitors. In our context, however, suppliers have no opportunity to shop for auditors because the brand selects them. As discussed above, the motivation of the firm selecting the auditors is also likely to be a boundary condition on the positive relationship between competition and quality. Competition is likely to raise auditor performance only if the firm selecting auditors wants rigor; otherwise, it is likely to undermine auditor performance.

Fourth, our finding that both insourced and outsourced auditors perform worse when they follow visits by colleagues from their own firm is an important contribution to the literature on audit rotation. This literature tends to focus on the rotation of *individual* auditors, including audit partners and lead auditors (e.g., Lennox et al., 2014; Short, Toffel, and Hugill, 2016), but does not address rotating audit *firms*, which is our focus. The few studies of audit firm rotation have several limitations that our study overcomes. First, many studies framed in terms of audit firm rotation do not, in fact, study it. Instead, they study the relationship between longer audit firm tenure and audit quality and extrapolate to form conclusions on the value of audit firm rotation (e.g., Al-Thuneibat, Al Issa, and Baker, 2011; Carcello and Nagy, 2004; Tepalagul and Lin, 2015). Second, studies that do examine audit firm rotation do so in contexts that have *mandated* financial audit firm rotation (Italy, South Korea, and Spain) and thus "may not generalize to the rest of the world due to the unique institutional features of these countries" (Lennox, 2014: 102). Specifically, Italy and South Korea bundled their audit firm rotation policies with mandatory auditor *retention* regulations and Spain "never actually enforced" its audit firm rotation policy before abandoning it. In an extensive review of the audit rotation literature, Lennox (2014: 103) concludes that "we still do not have a clear idea as to whether mandatory audit firm rotation would make audit quality better or worse" and that, beyond the effects of rotating lead auditors, we still do not know "whether audit firms *also* should be periodically changed." Our study fills this gap by controlling for rotation of lead auditors and measuring the effects of audit firm rotation on auditor performance.

Fifth, and more broadly, our findings suggest the need to revisit conceptions of legitimacy in auditing. Many have observed that the function of auditor independence is to confer legitimacy on the auditing process and its results (Boiral and Gendron, 2011; Kouakou, Boiral, and Gendron, 2013; Markell and Glicksman, 2014). Independence is a key signal of legitimacy for outside stakeholders who have little other information by which to judge audit quality (Darnall and Vazquez-Brust, 2018; Jia, 2018). But it is not clear that this visible symbol of legitimacy accurately reflects quality in the context of auditing business partners. Critical scholarship in the auditing literature suggests that there is a tension between auditing as a tool for information gathering and auditing as an accountability mechanism—or, put another way, between reliability and external perceptions of legitimacy (e.g., Power, 1997). Our finding that second-party auditors outperform third-party auditors in our empirical context lends some support to critics' charge that third-party audits are symbolic and performative tools that convey legitimacy to outside stakeholders, rather than rational tools that reduce monitoring costs and information asymmetry (Power, 1997). Scholars, advocates, and managers who seek to design high-quality assurance regimes should carefully consider what role independence can play in promoting reliable audits, what its limitations are, and what other tools are available.

Finally, our study contains insights that can inform the strategy, management, and operations literatures on sourcing by evaluating the performance of insourced, outsourced, and concurrently sourced production. This research focuses almost exclusively on explaining when firms will select these sourcing strategies rather than evaluating how these choices affect the quality of supplier performance. Thus, we respond to repeated calls for more attention to the outcomes of outsourcing decisions (Jia, 2018; Macher and Richman, 2008), particularly in the understudied context of outsourcing services (rather than production of tangible goods) (Jia, 2018).

Managerial implications

Our results have several important implications for managers monitoring their business partners' performance. First, our study challenges the prevailing wisdom that third-party audits are always the most reliable form of assessment. While such audits may be perceived as more legitimate by outside stakeholders, managers are likely to learn more from their own in-house audit team about what is happening on the ground at suppliers. Managers should periodically evaluate the effectiveness of their monitoring programs and work closely with third-parties to close any performance gaps detected. Managers who elect to perform audits in-house to gain information benefits might need to engage with stakeholders such as activists and NGOs, who often expect or demand third-party audits, about the reasons for their audit sourcing decisions and the benefits of different sourcing arrangements. Second, our findings help managers mitigate the tradeoffs they face when second-party audits yield higher-quality information than third-party audits but are more expensive. We demonstrate that managers need not sacrifice the typical cost-savings benefits of outsourced audits to bolster auditor performance. Integrating a concurrent sourcing strategy into an outsourced audit program can significantly improve the quality of outsourced audits. This finding is important for managers who wish to capture the reliability benefits of second-party audits but lack the resources or auditor availability to insource all audits. Finally, managers should consider rotating not only lead auditors but also auditing firms (including their own) to ensure that their auditors do not fall into the rut of organizational routines that may cause them to overlook important information.

Limitations and future research

We acknowledge several limitations to our study. First, our data come from a single firm that uses second-party auditors and several third-party auditors, which raises questions about the generalizability of our finding that second-party auditors outperform third-party auditors. It has proven difficult for apparel MNCs to converge on a common audit protocol and to share audit results with each other, so there are no available datasets that include audits conducted for multiple buyers by second- and third-party auditors. We use the best data available (obtained via a data-sharing agreement) to test our hypotheses and we encourage others to test both the generalizability of our findings and variations in the conditions under which they hold.

Second, despite being able to match on many variables that we believe are important to auditor assignment and performance, our reliance on observational data means that nonrandom auditor assignment, including the endogenous choice of concurrent sourcing, might occur along unobserved dimensions that could bias our results. Our interviews with MNC audit schedulers have not suggested that unobserved variables affect auditor assignment (and performance); nonetheless, the possibility exists.

Third, the proprietary audit dataset we obtained does not include variables to enable us to control for auditor training or professional credentials, which have been shown to affect audit results (Short, Toffel, and Hugill, 2016). While our MNC's second-party and third-party auditors received comparable training on the MNC's code-of-conduct requirements and while our direct observations of audits did not expose differences in audit skill, we cannot rule out the possibility that training and skill affected audit results in ways we cannot observe. We note that the fact that the MNC selects and pays all its third-party auditors, rather than allowing suppliers to do so, eliminates a significant potential source of auditor bias.

Finally, given the limitations of our data, we can only theorize—but not empirically test—the underlying mechanisms driving auditor performance. For instance, we cannot discern the extent to which the performance gaps we document are due to less effort being exerted by third-party auditors or to their inferior access to information. It is possible that suppliers are less transparent with third-party auditors because such auditors might intentionally or inadvertently disclose information provided in a third-party audit to other client brands. On the other hand, it is possible that suppliers are less transparent with second-party auditors because they are closer to the brand's employees who make sourcing decisions.

Our findings create opportunities for further study. Future research should investigate what combinations of independence and control promote effective monitoring of other types of business partners—such as franchisees, distributors, vendors, and purchasing agents—and whether these mechanisms can improve the performance of other service suppliers besides auditors. Studies could also investigate the efficacy of different strategies to close performance gaps between second- and third-party auditors and whether strategies that improve the performance of third-party auditors, such as concurrent sourcing, likewise improve the performance of second-party auditors. Future research could drill down into the mechanisms we theorize; for instance, testing auditor performance under varying reputation-risk conditions, such as reputation-damaging events or mandatory disclosure of audit results. Specifically, future studies could test the underlying theorized mechanism in H1b in the following way. First, they could collect reputationally damaging and supply-chain related news articles for brands that use insourced and outsourced auditors (and for brands they also have audit data for). Second, they could perform a differences-in-differences analysis to test whether the performance response of second-party auditors from the news release is both positive (resulting in more recorded violations) and significantly larger than that of third-party auditors.

CONCLUSION

In a world of complex, multi-layered business relationships and intricate business, regulatory, and social demands, firms must vigilantly monitor transaction partners. While monitor independence is often presumed to be the principal determinant of performance, we demonstrate that sourcing decisions, such as insourcing or concurrent sourcing, also play a role. We theorize these performance differences by synthesizing literatures on auditor independence and outsourcing decisions. Our findings contribute significantly to these literatures and provide important managerial insights for designing and implementing more effective business partner monitoring strategies.

The Impact of Remote Work on Service Quality: Insights from Quality Management Standards Inspections

Ashley Palmarozzo

Michael W. Toffel

INTRODUCTION

Remote work—where employees operate somewhere other than where work is typically conducted—has become increasingly common. Remote work is increasing in a broad array of industries, such as healthcare telemedicine, education, information technology, and financial services (U.S. Bureau of Labor Statistics 2022). Accelerated by the COVID-19 pandemic, many expect remote work to endure and perhaps constitute 20% of workdays in the post-pandemic U.S. economy (Barrero et al. 2021).

While the impact of remote work on job satisfaction, costs, and productivity have been studied, less understood is how remote work affects the quality of work being done. The few studies that have focused on quality have compared situations where the worker's location changed but their work process remained unchanged, such as when Bloom et al. (2015) found no quality difference between customer service representatives fielding customer calls while working from home versus at a call center office.

We build on this work by examining the quality implications of remote work when working remotely changes the way work is conducted. Specifically, we examine a service context that traditionally involved in-person visits during which auditors conduct face-to-face interviews, make direct observations during site tours, and review physical documents, but when conducting audits remotely these interactions are mediated by technology, including video conferencing and document sharing applications. This shift is part of a longer tradition of migrating in-person service operations such as bank tellers, in-store customer service desks, and medical consultations transitioning to remote offerings where workers interact remotely with customers via telephone, video, and online chats rather than face-to-face. These transitions, like the one we study, insert physical separation between the service provider and customer that, for reasons described below, can affect the quality of the work performed. We study how the shift from in-person to remote work affects quality of audits conducted to determine the extent to which a site has implemented all of the procedures required by one of six popular management system standards.

We theorize that remote auditing will produce lower quality (less comprehensive) audits by limiting an auditor's ability to gather sufficient evidence of violations, which results in auditors documenting fewer violations than are truly present at the audited site.[34] We theorize that this occurs through two mechanisms. First, auditors face greater difficulty accessing information at the audited site because auditors often rely on information that is physically embedded in an audited site to gather evidence of violations. While auditors typically tour sites in-person to gather evidence of violations, and video-mediated site tour in remote audits does not

[34] This approach has been used by others in contexts ranging from supplier codes of conduct (Short, Toffel, and Hugill 2016), restaurant inspections (Ibanez and Toffel 2020), and government health and safety inspections (Braithwaite and Makkai 1991; Gray and Shadbegian 2005).

convey all contextual information of the audited site (such as sound, smell, and cleanliness). Second, we theorize remote auditing will produce lower quality audits because auditors face greater difficulty accessing information held by other audit team members. Multi-auditor teams frequently communicate throughout an audit to exchange information, coordinate individual activities, and assess audit progress. We theorize that when auditors lack co-location (as in remote auditing) it reduces the likelihood that information is fully exchanged among team members so that they can coordinate their next activities and work together to identify possible areas of non-compliance.

We also hypothesize that two factors will moderate the audit quality disparity between audits conducted remotely versus in-person by mitigating or exacerbating the effect of one of the above mechanisms. First, we theorize that audits conducted by multi-auditor teams (versus audits conducted by a single auditor), which require coordination, will worsen the information access challenges among remote audit team members and thus exacerbate the quality decrement of remote auditing. Second, we theorize that information access challenges with remotely audited sites will attenuate (and thus the quality of remote audits will lessen) when auditors have greater in-person exposure to the audited site. This is because such auditors are more able to rely on experience-based heuristics that can prompt questions to uncover problems given remote auditing lacks access to onsite cues.

To test our hypotheses, we obtained data from a large company that conducts audits around the world to assess sites' compliance with various management system standards. Specifically, we examine their 35,000 audits conducted in-person or remotely from 2019-2022 pertaining to six of the world's most popular management system standards: the Quality Management Systems (ISO 9001), Environmental Management Systems (ISO 14001), Information Technology Management Systems (ISO 27001), Occupational Health and Safety Management Systems (ISO 45001/OHSAS 18001), and Medical Devices Quality Management Systems (ISO 13485). 61% of these audits were conducted fully in-person and the remaining 39% were conducted fully remote. The largest proportion of these audits were conducted in the United Kingdom, China, India, and the US, with the rest spread across 100 countries.

Our empirical analysis indicates that remote audits yielded 0.40 (or 25%) fewer violations than in-person audits, which is consistent with our theory that remote audits result in lower quality audits than in-person audits. We find that this quality gap is exacerbated when audits are conducted by multi-auditor teams (versus those conducted by a single auditor), which supports the presences of one mechanism underlying our theory, that remote auditors in multi-auditor teams face challenges accessing information from one another. We also find evidence that teams with greater exposure to the audited site attenuate the quality gap between remote and in-person audits.

As an extension, we also tested for the presence of the first mechanism underlying our theory: that remote auditors face an increased burden to obtain information from audited sites necessary to identify violations. We do so by exploiting the fact that auditors typically use different auditing techniques to identify violations of particular management system standard elements, and the reduction in information that remote auditors are able to obtain when assessing these elements varies by auditing technique. Specifically, we examine whether the quality deficit of remote audits is especially pronounced for those standard elements where auditors tend to identify problems via direct observation (where remote auditors are at a particular disadvantage compared to in-person auditors) versus those standard elements where violations are typically identified using document review. Consistent with our theory, we find that negative effect of remote audits on audit quality is significantly larger for violations of standard elements typically identified via direct observation relative to those typically identified via reviewing site documents.

Our results make several contributions to two streams of literature. Our work contributes to the remote work literature by finding evidence that remote work can impede work quality. When viewed in light of prior research that found no quality difference associated with remote work conducted by customer service representatives or patent examiners (Bloom et al. 2015; Choudhury et al. 2021), our results suggest that the quality implications of remote work depend on the extent to which working remotely changes work processes. In particular, remotely conducting work that has traditionally required gathering information in-person, such as conducting medical exams and conducting various types of inspections, and should trigger special efforts to safeguard against quality loss. (Bloom et al. 2015; Sun et al. 2020; Bettinger et al. 2017). Our research also contributes to the literature on monitoring quality. While there is significant scholarship on audit-, auditor-, and institutional-level factors that affect audit quality for in-person audits, our study is the first to our knowledge to examine whether varying the audit format affects quality. Understanding the relationship between audit format and quality is especially timely given the growing interest in remote auditing in regulatory, supply chain, and standard certification contexts (U.S. Department of Health and Human Services, 2022; IAF, ILAC, and ISO 2021).

Our findings also have clear managerial implications, offering guidance to those seeking or providing remote auditing services, whether of management system standards, franchisee compliance with franchiser terms, or establishments' adherence to supplier codes of conduct. Given that remote auditing is a useful audit format because it reduces auditor travel time and associated travel costs and increases auditing flexibility, our results can guide audit providers on how to staff audits to minimize quality concerns associated with remote auditing, such as by focusing the use of remote audits on smaller sites that can be audited by one auditor. Moreover, audit providers can use our results to develop a hybrid auditing approach – for example, by performing documents review activities remotely and performing observational activities onsite - which is described in further detail in the Discussion section below.

RELATED LITERATURE

Our work relates to the literature on remote work that compares work done remotely to in-person, and to the literature that examines the effectiveness of monitoring the management practices of companies' business partners.

Remote Work

The literature on remote work has focused on the performance implications of several forms of remote work. One form refers to employees working remotely from their workplace, such as working from home and working from anywhere. Allowing workers to work remotely has been found to lead to higher worker productivity, more job satisfaction, and lower attrition (Bloom et al. 2015; Choudhury et al. 2021).[35] Sun et al. (2020) found that, in a hospital emergency room setting, adopting telemedicine for initial patient screening improved the hospital's operational efficiency through more flexible physician allocation. Another stream of research has examined the performance implications of organizations offering online service to supplement their conventional in-person service. For example, banks and retail firms that adopted online service channels saw increased customer retention (Cgulgulampbell and Frei

[35] Tan and Netessine (2020) also find that introducing tabletop technology in restaurants to assist in customer food orders increased waiter productivity and sales per table.

2010, Xue et al. 2011, Buell et al. 2010), greater use of *offline* service channels (Campbell and Frei 2010), and more overall customer interactions (Campbell and Frei 2010, Xue et al. 2011, Bell et al. 2018). In the healthcare industry, once patients were provided with an online channel to communicate with doctors, those patients made more *in-person* doctor appointments (Bavafa et al. 2018, Rajan et al. 2019). One study that examined the reverse transition— supplementing their conventional online service by adding in-person service—found that sales conversions increased and product returns decreased, improvements that resulted from changes in customer behavior (Bell et al. 2018).

Closer to our research are the few studies that have examined how working remotely affects the quality of work being accomplished, which have yielded mixed results. For example, two studies have found no evidence that work quality differed between remote and in-person workers, whether comparing customer service representatives who worked from home versus at a call center (Bloom et al. 2015) or patent examiners working from home versus "working from anywhere" (Choudhury et al. 2021). In healthcare, Sun et al. (2020) found that physicians remotely delivering health care services to patients in hospital emergency departments improved one dimension of service quality—reducing patient length-of-stay, both by reducing patient waiting time (treatment delays) and by more flexibly allocating the (remote) physicians to patients—with no erosion in care quality (readmission rates and in-hospital mortality rates). Another study found better patient health outcomes resulted when doctors supplemented in-person visits with virtual channels to communicate with patients (Bavafa et al. 2018). In the education context, Bettinger et al. (2017) found worse educational quality outcomes when students took a course online compared to those who took the course in-person, theorizing that difference was due to the online students being subjected to less instructor oversight and thus being less motivated.

Our study also examines the quality implications of remote work, but differs along two key dimensions. First, prior studies in this domain theorize that differences in the quality of work performed remotely versus in-person result from differences in service delays (Sun et al. 2020) and oversight (Bettinger et al. 2017), whereas these dimensions do not meaningfully differ between remote versus in-person work in the auditing context we study.

Second, prior studies that compare the quality of office work performed remotely to in-person consider scenarios in which worker locations differed but where the work did not entail in-person customer interaction. For example, when Bloom et al. (2015) compared the work of those working from their office to those working remotely from home, all of these workers were conducting sales calls via telephone. Similarly, Choudhury et al. (2021) compared remote patent examiners working from home to those working from anywhere, but in both cases the work entailed examining documents. In contrast to these settings, we compare work (auditing) conducted in-person at customer sites in the presence of customers to work conducted remotely where auditors instead worked from their office or home and engaged with customer sites via audio/video technology. Our setting therefore enables us to compare a particular form of remote work—where workers and customers are not co-located—to the in-person scenario where they are co-located. This distinction provides greater variation in coordination and information flow that we theorize will affect work quality. Monitoring Business Partners' Management Practices

Our work also relates to the literature that examines the effectiveness of companies monitoring their suppliers' and other business partners' management practices, often as part of their due diligence efforts to ensure they meet minimal standards. Some buyers, for example, conduct monitoring to mitigate the risk of incurring negative reputation spillovers, such as from media reports that a food retailer chain customers' were sickened by contaminated supplies or an apparel brand's suppliers employed child labor. Our work is most closely related to those studies that identify determinants of monitoring quality, where worse quality refers to problems being underreported in audit reports (e.g., Gul, Wu, and Yang 2013, Short, Toffel, and Hugill 2016). Such monitoring or audit quality problems have largely been attributed to characteristics of individual auditors and audit firms.

In terms of individual auditors, research has found that lower quality audits result from less-trained and less-experienced auditors (Macher, Mayo, and Nickerson 2011, Short, Toffel, and Hugill 2016), and when individual auditors have prior experience auditing the business partner (Short, Toffel, and Hugill 2016, Ball, Siemsen, and Shah 2017). Studies have also found audits to be of lower quality when conducted by all male audit teams (Short, Toffel, and Hugill 2016), and when they are conducted later in an auditor's workday, likely owing to fatigue (Ibanez and Toffel 2020).

Audit company characteristics have also been shown to affect audit quality. For example, several studies have found evidence consistent with economic conflicts of interest, including worse quality audits when monitoring firms are (a) paid directly by the monitored entity (Duflo et al. 2013, Short, Toffel, and Hugill 2016), or (b) have opportunities to cross-sell other services to those they are monitoring (Pierce and Toffel 2013). Moreover, worse quality audits are performed by companies that face greater competition (Bennett et al. 2013), and when audits are conducted by auditors from the same firm that that performed the prior audit even when the individual auditors are new to the site (Ibanez et al. 2022).

These studies have identified a wide range of individual auditor and audit firm-level characteristics that influence audit quality, but they all focus on audits performed in-person onsite. In contrast, we examine whether and how audit quality varies between audits performed in-person and those performed remotely, and how such differences are moderated by various attributes of audit teams and individual auditors.

HYPOTHESES

Management system standard certification audits are conducted to assess whether a business complies with best management principles and practices set forth in various management systems standards covering a range of areas, such as quality management (ISO 9001), environmental management (ISO 14001), and occupational health and safety (ISO 45001/OHSAS 18001). Some companies voluntarily seek certification to one or more standards for a variety of reasons, such as to improve the efficiency of their business, to continuously improve the quality of their products or services, and/or to signal their quality to other stakeholders. Each standard has several clauses and sub-clauses, which address different areas of best practices.[36]

A site's compliance with a management system standard is assessed by auditors using through three primary auditing techniques: reviewing documents (such as policies, procedures, and training records), interviewing employees and managers, and direct observation (ISO 19011: Section 6.4.7). Direct observation refers to auditors touring the site to observe processes and work activities. Auditors rely on these techniques to collect evidence, which they then compare

[36] For example, ISO 9001's standard has different clauses for best practices for leadership, operations, and performance evaluation.

to all elements of the management system's requirements in order to determine whether compliance with each element is met or if any non-conformities (violations) exist.

For all audits, we consider an audit resulting in more violations to reflect higher audit quality, all else equal. Auditors collect evidence to prove the existence of both compliance and non-compliance. However, if auditors cannot collect sufficient evidence to prove compliance (which would otherwise lead to a violation), the audited site is given several opportunities to furnish the evidence to confirm compliance. Specifically, management system auditing rules require auditors to review their preliminary findings of noncompliance "with the auditee in order to obtain acknowledgement that the audit evidence is accurate and that the nonconformities are understood. Every attempt should be made to resolve any diverging opinions concerning the audit evidence or findings" (ISO 19011: section 6.4.8). Audited sites are thereby provided the opportunity to share additional evidence to rectify erroneous preliminary results indicating non-compliance, mitigating the risk that audit reports will report false violations. Audited sites are incentivized to provide such evidence if it exists because it would otherwise result in a violation, which requires audited sites to invest time (and potentially money) to remediate. Therefore, a reduction in audit quality means auditors will underreport violations that are actually present at the audited site rather than report fictitious violations.

While we test our hypotheses using the empirical context of auditing adherence to management system standards, our theory applies to other forms of remote work in which working remotely affects how work is conducted. This includes settings where direct observation is used to gather information, such as during medical exams, supply chain audits, franchisee monitoring, and regulatory inspections of establishments (e.g., occupational safety, environment, hygiene), our results should trigger special effort by those engaged in working remotely health and safety regulatory inspections. The transition from in-person to remotely auditing management system standards is representative of a much larger shift to remote service provision that is occurring in a wide array of industries. Consider banking where many have transitioned from in-person teller service to online banking, healthcare where some in-person doctor appointments are shifting to telemedicine, customer service departments and the rise of call centers, and education where some in-person teaching is being replaced with online courses. Industries ranging from banking to medicine have adopted remote services to reduce costs, avoid customer travel, reduce wait times, and to increase the hours of service provision , and to increase worker productivity. While remote auditing was primarily adopted to overcome travel restrictions prompted by the COVID-19 pandemic, the reduction in auditor travel time and associated costs, the additional flexibility of scheduling remote audits, and the additional productivity of remote auditors suggest that remote auditing is likely to become an enduring management practice.

Remote Audits and Audit Quality

While remote auditing can reduce the cost of auditing by eliminating auditors' travel-related time and expenses, we argue that remote auditing will produce less comprehensive—and thus lower quality—audits, because it reduces the availability of information that auditors need in order to identify all violations present at an audited site.

Auditors often rely on contextual knowledge in an audited site gained through observation to inform their assessment. Prior literature demonstrates that the observing of contextual attributes such as equipment, procedures, workers, and work setting, can improve problem-solving efforts by broadening the set of information that the problem-solver has access to (Tyre and von Hippel 1997). In our context, auditors may use information gained through direct observation to assess whether a site is complying with worker safety, or environmental, areas of a standard. However, the technology mediation required by remote audits can constrain the quality and amount of information auditors can access, which can lead remote auditors to miss contextual knowledge that they would otherwise observe during an in-person audit. Similarly, because technology mediated interpersonal interactions tend to reduce participant engagement and verbal cues compared to in-person interactions (Bohannon et al. 2013, Reid and Reid 2005), the technology mediated interviews conducted during remote audits are likely to convey less information than in-person interviews conducted in the course of in-person audits. As such, less evidence may be collected in remote audits which could otherwise lead to finding more areas of non-compliance and thus more violations.

The extent to which additional effort is required to identify and record violations in remote audits compared to in-person audits will likely vary by the audit activity used to detect specific violations. Whereas the process of reviewing documents remotely or in-person is quite similar, this is not so for those management standard elements that require auditors to physically observe in order to assess compliance. In contrast to in-person auditors touring the site to uses their senses of sight, sound, and smell to assess conditions, remote auditing requires the use of video-mediated experience, where the auditor directs a worker to walkthrough the site. This reduces the amount of information that remote auditors can access and learn from.

We therefore hypothesize:

> *H1: Remote auditing yields worse quality audits (fewer violations) than in-person auditing.*

H2: The decrement in audit quality associated with remote auditing is exacerbated for audited areas where compliance is assessed via physical observations rather than via document review.

Next, we hypothesis two factors that will moderate the relationship between audit format and quality.

Coordination Among Multi-Auditor Teams

The adoption of remote auditing not only creates physical separation between the auditor(s) and the audited entity, it also creates physical separation between audit team members. Just as we theorize that physical separation between the audit team and audited entity reduces information flow between them, we also theorize that reductions to information flow will befall geographically distributed audit team members ("virtual teams"), which will further reduce the quality of remote audits.

Prior literature has documented several challenges virtual teams face, many of which can result from coordination problems. Cramton (2001) proposes that coordination problems arise in virtual teams because it is more difficult to obtain mutual knowledge among team members, which is when all team members have identical information and all team members know that they have identical information. Uneven information distribution among team members, or a lack of mutual knowledge among team members, reduces the ability of team members to identify and resolve problems, and ultimately impedes their performance. Virtual teams are more likely to report increased levels of team conflict, possibly because they lack mutual knowledge that they would otherwise use to resolve potential issues early on in in-person settings (Hinds and Mortensen 2005). Virtual teams are also more likely to experience lower reported levels of team trust (Pennaroja et al. 2013) and psychological safety (Gibson and Gibs 2006), both of which can affect the team's ability to openly communicate and resolve potential problems with team members.

In our setting, audits are staffed by one or more auditor, with multi-auditor teams requiring frequent communication with one another to exchange information (which ensures mutual knowledge across team members), coordinate their individual activities (and possibly reassign work among team members), and assess audit progress (ISO 19011: section 6.4.4). Coordination among multiple auditors is especially important because an auditor's next activity could be dependent on the result of a different auditor's prior activity. We argue that when remote audits create virtual multi-auditor teams, it reduces the likelihood that information is fully exchanged among team members so that they can coordinate their next activities and work together to identify possible areas of non-compliance. However, the chance of uneven information exchange is likely non-existent for single-auditor teams because there are no other individuals to coordinate information across. Multi-auditor teams' inferior ability to coordinate should then exacerbate the theorized negative effect of remote audit format on audit quality.

We therefore hypothesize:

H3: The decrement in audit quality associated with remote auditing is exacerbated when audits are conducted by multi-auditor teams.

Auditor Exposure to the Audited Site

Prior research on in-person supplier auditing and regulatory inspections has found that auditors assigned to audit sites that have previously audited tend to conduct less comprehensive audits (Short, Toffel, and Hugill 2016; Ball, Siemsen, and Shah, 2017). One potential mechanism driving this effect is “bounded awareness” (Chugh and Bazerman, 2007), a cognitive bias that in our context would lead such repeat auditors to overly focus on problem areas they have previously identified at that site. According to Kumar and Chakrabarti (2012), people tend to rely on information that agrees with tacit knowledge gained through prior experience, which can limit their ability to recognize new problems. In our context, this means that repeat auditors are more likely to make decisions based on tacit knowledge gained through their prior exposure with the audited site instead of using information that is easily available in the present audit to guide their decisions.

Whereas returning to a site to conduct another in-person audit tends to impede the quality of such audits, we hypothesize that auditors who have more prior in-person audits at the audited site will conduct higher quality remote audits of that site. Having more prior in-person auditing experience at the site can help overcome the information access challenges associated with remote audits by equipping remote auditors with more accumulated tacit knowledge of the site gained through their prior in-person audited site visits. Therefore:

H4: The decrement in audit quality associated with remote auditing is attenuated when auditors have more prior in-person auditing exposure to the site.

DATA AND MEASURES

Empirical Context

We obtained data from a global company that provides auditing, certification, training, and consulting to organizations around the world, which has requested anonymity as a condition of sharing data with us. The company’s auditing and certification practices are accredited by many leading accreditation bodies, which provides third-party assurance that their auditing practices meet international auditing standards.

This company shared all 438,084 certification audits that the company conducted from January 1, 2019 to October 7, 2021 pertaining to six of the world's most widely used management system standards, including the management of quality (ISO 9001), environment (ISO 14001), information security (ISO 27001), occupational health and safety (OHSAS 18001 and ISO 45001), and medical device quality (ISO 13485). In the course of obtaining and retaining certification to any of these standards, sites receive an *initial audit* to assess whether its management system meets the requirements of given standard, and, if so, the company issues a three-year certification.[37] Certified sites are subsequently subjected to *surveillance audits*, typically one in each of the following years, which provide an ongoing assessments of the site's adherence to the standard. Auditors follow the same audit process in surveillance audits, but check a subset of areas within the standard.[38] In the third year, typically just before the certification term expires, a *re-certification audit* is conducted to provide another comprehensive assessment of the site's adherence to the standard, and a successful audit results in a three-year renewal of the site's certification. In our analysis, we focus only on surveillance audits due their numerosity in the data and due to statistical power: our dataset includes multiple surveillance audits (and in many cases, some conducted in-person and others remotely) for a given site being audited to a particular standard.[39]

[37] The initial audit is performed in two stages. In stage 1, auditors check whether the necessary procedures are developed to meet the quality benchmark set forth in the standard. Stage 2 assesses whether those procedures were implemented and assesses the efficacy of the procedure implementation. Failure to implement those procedures would result in a non-issuance of the certification until the site provides evidence that the failures have been rectified.

[38] All areas of the standard are audited across the two surveillance audits conducted over consecutive years during the three-year certification period.

[39] Because initial audits are typically conducted once per site-standard, we cannot compare in-person to remotely

Our raw data includes 125,727 surveillance audits of 48,956 sites corresponding to the six management system standards mentioned above during the nearly three-year period of our sample. Some sites adopted more than one standard, and so our dataset includes 64,222 "site-standards," where each site-standard refers to a unique pair of one site and one management system standard being audited (referred henceforth as a site-standard dyad). Our data tends to contain several surveillance audits for each site-standard.

Because our specification (described below) includes site-standard dyad fixed effects to enable comparisons of in-person audits to remote audits of the same site for a particular standard, our model requires at least two audits per site-standard. This excludes 24,308 audits for which our data included only one surveillance audit for a given site-standard. We then omit 5,197 audits of 1,080 sites (1,576 site-standards) with data anomalies that are either data entry errors or that do not reflect the typical course of audits. We next omit 28,408 audits that received two or more surveillance audits in a given year because such audits tend to have a smaller scope than annual surveillance audits, and their inclusion in the sample could introduce bias into our analysis (e.g., if unobserved differences in such "mini-audits" are correlated with our hypothesized variables and the number of violations).[40] 4,427 of these omitted audits are hybrid audits, where a portion of the audit is performed in-person and the rest remote.[41] Finally, our Poisson regression

---

conducted initial audits of a site-standard. Similarly, re-certification audits are conducted roughly every three years, and we have very few instances in our dataset of variation of in-person versus remotely conducted re-certification audits of a site-standard. Thus, we exclude initial audits and re-certification audits from our analysis and instead focus on surveillance audits.

[40] We considered combining these multiple surveillance audits conducted in a calendar year for a single site-standard, thinking they might simply represent a surveillance audit split into parts. But our preliminary analysis led us to conclude that aggregated surveillance audits are not comparable to typical once-per-year surveillance audits: the former required substantially more auditor-days than the latter, even after accounting for sites' employment level and industry, the factors that audit standards specify was determining the number of auditor-days to be allocated to each audit. As such, we dropped all site-standards that included multiple surveillance audits being conducted in a calendar year.

[41] 573 hybrid audits are same-date hybrid audits, where the same audit taking place on the same day is recorded twice in the data, except one instance of the audit is in-person and the other instance is remote. 3,854 hybrid audits

framework omits audits with all zero outcomes within a site-standard, which omits an additional 35,247 audits. More information on sample selection can be found in Appendix Table B-1.

This results in a sample of 35,247 surveillance audits of 16,986 site-standards to which we apply our data analysis. Of these, 54% surveillance audits assessed adherence to for the ISO 9001 quality management standard, 18% to the ISO 14001 environmental management standard, 12% to the ISO 27001information security management standard, 11% to the OHSAS 18001 or ISO 45001 occupational health and safety management standards, and 5% to the ISO 13485 medical device quality management standard. These surveillance audits were conducted for sites located in 109 countries, and the largest share of audits were for sites in the United Kingdom (35% of sites), China (9%), India (8%), and the United States (7%).

Auditors assess a site's adherence to each of these standards by reviewing the site's documents, interviewing its employees, and touring its facility. Auditors have traditionally performed these tasks on-site at the audited site, but a growing number of audits have been conducted remotely, and the use of remote audits accelerated during the COVID-19 pandemic to reduce exposure and travel.[42] Remote audits are performed off-site from the auditor's home or office, and rely on audio/video technology. Remote audits and in-person audits are intended to entail the same level of scrutiny and thus produce the same quality of audits.

Dependent variables

are multi-date hybrid audits, where two separate audits are recorded at two different points in time, one being remote and the other in-person, within the same calendar year.

[42] Starting March, 2020, an audit's format (remote or in-person) is determined by performing a risk assessment to assess whether an auditor can safely access an audited site in-person or whether a site can be successfully audited remotely.

We measured the degree to which a site does not fully comply with all clauses of a management system standard as the number of violations (*violations*) discovered and recorded in a surveillance audit. This approach has been used by others in contexts ranging from supplier codes of conduct (Short, Toffel, and Hugill 2016), restaurant inspections (Ibanez and Toffel 2020), and government health and safety inspections (Braithwaite and Makkai 1991; Gray and Shadbegian 2005).[43]

Audits can result in violations of different aspects of management system standard requirements. For example, violations identified during an audit of the ISO 9001 Quality Management System Standard could correspond to the standard's clauses addressing the site's operations, leadership, or personnel that do not fully meet a particular aspect of the standard. An violation might be recorded, for example, if the site has not established clear communication channels with clients or customers or has not devoted resources to establish proper quality assessment processes.

Discussions with the company that provided our audit data indicated that auditors tend to rely on one of several auditing activities—document review, direct observation, and interviews—to identify non-compliance with each clause within a standard. For example, the ISO 9001 Quality Management System Standard's clause 6.2 requires organizations to maintain documented information about their management system quality objectives and how to achieve them, which auditors tend verify by reviewing those documents. In contrast, ISO 9001 clause 8.5.1 requires that suitable infrastructure and environment be used for site operations, which

[43] Each violation is categorized as major or minor. A violation is classified as major when it impedes a management system's capability to achieve a site's objectives (e.g., financial success or product quality), and is otherwise classified as minor. In our sample, 97.5% of violations are minor and the remaining 2.5% are major. Our results described below are robust to the exclusion of major violations as seen in Appendix Table D-3.

auditors tend to verify by directly observing the site's production facility. To test hypothesis 2, we also categorized each audit's *violations* into the number of violations associated with standard clauses where violations were primarily discovered by document review, interviews, and direct observation based on information we obtained from the data provider Of all the clauses in the six standards we study, 301 are primarily discovered by document review, 61 by direct observation, 44 by interviews, and 48 where there is no single primary activity (e.g., two were equally relied upon). We focus on violations of the former two sets of clauses, which we refer to as violations from document-review detection mode and violations from observation detection mode. Doing so avoids concerns of limited statistical power of violations corresponding to interview-based clauses given how rare they are, and avoids the clauses where a single primary auditing technique was not indicated.

Independent and moderator variables

We coded *remote audit* as 1 when an audit was conducted remotely, and 0 when it was conducted in-person. We coded *multi-member team* as a dummy variable coded 1 when an audit was conducted by more than one auditor, and 0 when an audit was conducted by one auditor.[44]

To measure the audit team's in-person familiarity with the audited site, we coded *average prior in-person site exposure (log)* as the sum of prior in-person site visits any audit team member has had prior to the focal audit divided by the audit team size, which we log to reduce skew.[45]

---

[44] Of the 20% of audits staffed with more than one auditor, 76% had two person audit teams, 20% had three person audit teams, 3% had four person audit teams, and 1% had audit team sizes ranging from five to nine auditors.
[45] We use an average measure because the variable that included just the numerator was highly correlated with *multi-member team*.

Control variables

We control for a number of audit-, auditor-, and site-level variables that may be correlated with our outcome measure. We measured an audited site's complexity as the number of *staff-days (log)* scheduled to perform an audit. Staff-days are determined using International Accreditation Forum (IAF) documentation and are based on site size (employment) and site risk (measured by site industry). We use the natural log in our models to reduce skew. We constructed *COVID time period* as 1 for audits conducted from March 2020 through December 2020, and 0 otherwise.

We coded *prior remote site exposure* as 1 when at least one audit team member had previously conducted a remote audit of the site, and 0 none had done so.[46] We coded *focal standard advanced training* as a binary variable which equals 1 if any audit team member other than the team leader held an advanced qualification for the focal audit's standard, and 0 if no other team member holds the advanced qualification.[47] We also constructed *maximum auditing experience (log)* as the maximum number of prior audits any auditor on the team had conducted since the start of our sample period up until the start of the focal audit, which we log (after adding one to avoid introducing missing values associated with taking the natural log of zero) to reduce skew. We created *female on audit team,* which equals 1 if at least one audit team member is female, and 0 otherwise.[48] We constructed *percent outsourced* as the percentage of audit team members are not employees of the certifying organization.

[46] We chose this measure instead of measuring *average prior remote site visits (log)* because the binary variable captures 87% of variation in *average prior remote site visits (log)*.
[47] An auditor is considered to hold an advanced qualification for a given standard if they hold a "lead assessor" qualification for that standard. We code this variable to 0 for the 3,474 audits were this data was missing, and created a dummy variable coded 1 to flag those observations.
[48] This variable is missing for 11,324 audits and, when missing, is set to 0 and modeled with a binary variable equal to 1 when missing.

We coded *multi-standard audit* as 1 if an audit for a different standard takes place at the time of the focal audit, and 0 otherwise.

Table 2.1 provides summary statistics and Table 2.2 reports correlations.

Table 2.1: Summary Statistics

	N	Mean	SD	Min	Max
violations	35,247	1.42	1.65	0	58
Violations from document review detection mode[1]	35,247	1.02	1.38	0	56
violations from document review detection mode[a]	29,551	1.21	1.43	0	56
Violations from observation detection model[1]	35,247	0.15	0.46	0	9
violations from observation detection mode[a]	8,080	0.67	0.75	0	9
remote audit	35,247	0.39	0.49	0	1
multi-member audit team	35,247	0.20	0.40	0	1
average prior in-person site exposure	35,247	2.89	3.22	0	48
average prior in-person site exposure (log)	35,247	1.10	0.71	0	3.89
focal standard advanced training[b]	31,335	0.23	0.42	0	1
staff-days	35,247	1.56	1.07	0.5	30
staff-days (log)	35,247	0.29	0.53	-0.69	3.40
COVID time period	35,247	0.31	0.46	0	1
prior remote site exposure	35,247	0.15	0.35	0	1
maximum auditing experience	35,247	344.17	211.02	0	1048
maximum auditing experience (log)[c]	35,247	5.53	1.00	0	6.96
female on audit team[b]	22,131	0.24	0.43	0	1
percent outsourced	35,247	0.28	0.42	0	1
multi-standard audit	35,247	0.10	0.30	0	1
audit year	35,247	2019.91	0.79	2019	2021
audit sequence[d]	35,247	6.52	3.63	1	16

Notes:
[1] The means of these two variables do not sum to 1.42 (the mean of *violations*) because we omit ISO 13485 audits (we did not obtain a mapping of compliance items to detection mode) and because we omit violations from the interview detection mode.
[a] These variables have smaller Ns than that of the full data sample because the model that tests this hypothesis includes site-standard-detection mode fixed effects (as opposed to site-standard fixed effects included in the other models) and this difference omits additional observations when there are no violations recorded from a given detection mode within a site-standard. [b] denotes variables with some missing values (and thus smaller N reported here); in the regression models these missing values are recoded to zero and we include a dummy variable flagging those observations where such recoding occurred. [c] one is added to the un-logged versions of these variables before taking the natural log to avoid their logged values be missing. [d] denotes variables that are winsorized at the 95th percentile.

Table 2.2: Correlations

		(1)	(2)	(3)	(4)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
(1)	violations	1.00										
(2)	remote audit	-0.12	1.00									
(3)	multi-member team	0.02	-0.12	1.00								
(4)	average prior in-person site exposure (log)	-0.01	-0.34	0.06	1.00							
(6)	focal standard advanced training[a]	0.04	-0.08	0.65	0.02	1.00						
(7)	staff-days (log)	0.17	-0.05	0.48	0.02	0.42	1.00					
(8)	COVID time period	-0.06	0.38	0.00	-0.11	0.03	0.01	1.00				
(9)	prior remote site exposure	-0.05	0.28	-0.05	0.13	-0.07	-0.01	-0.07	1.00			
(10)	maximum auditing experience (log)	-0.10	0.03	0.19	0.32	0.05	-0.05	0.01	0.15	1.00		
(11)	female on audit team[a]	0.03	0.03	0.08	-0.03	0.07	0.03	0.02	0.01	0.04	1.00	
(12)	percent outsourced	-0.02	-0.05	0.08	0.05	0.00	0.09	-0.02	-0.04	-0.15	-0.22	1.00
(13)	multi-standard audit	-0.04	-0.06	0.22	0.11	0.19	-0.07	-0.01	-0.07	0.11	0.00	0.01

Notes: N = 35,247 audits. [a] denotes variables where missing values are recoded to zero.

EMPIRICAL MODELS AND RESULTS

Empirical specification

We estimate the effect of our hypothesized variables for all hypotheses using a Poisson pseudo-maximum likelihood regression that predicts the number of *violations* recorded in each audit. Each model we use to test hypotheses 1, 3, and 4 includes the following primary independent variables: *remote audit*, *multi-member team*, and *average prior in-person site exposure (log)*. The models used to test hypotheses 2-4 also included one interaction term that interacts *remote audit* with one of the other hypothesized moderate variables.

All models include several control variables. We include *staff-days (log)* because the more time an auditor has at the site, the more opportunities they have to uncover more violations. To account for the possibility that an auditor's advanced qualifications obtained for the focal audit's standard could affect audit qualify, we include *focal standard advanced training*. To ensure that an audit team's prior exposure remotely auditing a site does not confound our testing whether an audit team's prior in-person site exposure moderates the relationship between remote audits and audit quality (hypothesis 4), we include *prior remote site exposure*. Since cumulative auditing experience has been shown to influence audit outcomes (Macher, Mayo, and Nickerson 2011; Short, Toffel, and Hugill 2016; Ball, Siemsen, and Shah 2017), we include *maximum auditing experience (log)*. We include *female on audit team* as prior research has found that audit teams with at least one female find more violations (Short, Toffel, and Hugill, 2016). Because prior research suggests that using your own employees (rather than external contracted workers) increases the number of recorded violations (Ibanez et al. 2022), we include *percent outsourced*. We also include *multi-standard audit* to account for the possibility that sites' violation counts could be impacted by their being audited against several standards simultaneously.

All models also include audit year fixed effects to control for secular factors that might affect auditor stringency and/or violation counts. While year fixed effects can control for differences between years, the COVID-19 pandemic that began in March 2020 might be another temporal factor that affected audit results (e.g., because some sites operated with fewer workers), and so we include *COVID time period.* All models also include *audit sequence* fixed effects, which reflect how many audits a site has received from the audit company before the focal audit (1st audit, 2nd audit, etc.), winsorized at the 95th percentile (20th audit) to minimize the influence of outliers. We include audit sequence fixed effects to control for the possibility that sites become more compliant the more they are audited because they have had more prior audits to expose, and thus resolve, violations.

The models that test hypotheses 1, 3 and 4 include site-standard dyad fixed effects to control for time invariant factors (observed and unobserved) associated with each combination of site and management system standard that could affect compliance levels, such as its industry and location.

The model specification that tests hypothesis 2 is similar to our primary model, except that here we include a dummy variable *observation* and divide each audit observation into two observations that are otherwise identical except that in one observation, the dependent variable *violations* includes only *violations from observation detection mode* and *observation* is coded 1, and in the other observation, *violations* includes only violations from *document review detection mode* and *observation* is coded 0. We include the interaction term, *observation X remote audit*, to obtain the coefficient of interest. This model includes site-standard-detection mode fixed effects (e.g., Site A – ISO 9001 – document review) to similarly hold constant time invariant characteristics specific to a site, standard, and auditing detection mode that could affect compliance levels.[49] Our modeling specification that tests this hypothesis is akin to separately estimating the same model on two samples, one on the sample where violations are recorded primarily through observation (*observation* equals 1), and another sample where violations are recorded primarily through document review (*observation* equals 0), and then performing a Chow test to test for coefficient equality (Chow 1960). In estimating one model which combines both samples, this allows for an easier interpretation of results and use fewer degrees of freedom because we do not separately estimate coefficients for all other model variables. However, in doing so we assume no differential correlation between other control variables violations that stem from observation vs. document review detection modes.

Identification strategy

Remote audit assignment. In order for our results to provide an unbiased estimate of the effect of remote audits on violation counts, unobserved variables should not be correlated with both an audit's outcome and the decision to perform the audit remotely. Through conversations with the data provider, we learned that remote audits were first adopted in 2019 as a cost savings measure, and they became much more widely used as soon as the COVID-19 pandemic began, an exogenous event that is plausibly orthogonal to violation counts. Of the remote audits in our 35,247 estimation sample, 0.2% occurred between January 2019 to March 2020 (when the pandemic became widespread), 53.1% during the remainder of 2020, and 44.7% in 2021.

To determine if a particular audit would be remote or in-person, the auditing company used a two-step process. First, assuming country laws allowed auditors to access a site to conduct an in-person audit, the auditing company used their standardized "Pre-visit COVID-19 checklist"

[49] Due to our including this fixed effect, the main effect of *observation* is absorbed in the model and is thus not reported in any results.

to determine whether their auditors could safely conduct an in-person audit without incurring significant risk of COVID-19 exposure.[50] Second, the company assessed whether the audited site has the IT infrastructure to safely and rigorously deliver a remote audit. Any site that could not be visited in-person safely and did not have infrastructure available to rigorously deliver a remote audit would not receive an audit until it was safe to do so in-person. Moreover, audited sites would not receive a remote audit if the audited site did not agree to the remote audit format.

Auditor Assignment. There are three reasons why we believe the assignment of auditors to audits is plausibly exogenous to our hypothesized variables and violation counts. First, we learned through conversations with the audit company that auditors were assigned to audits (of any format) based on their availability and that for sites in some more complex industries such as aerospace and medical devices, auditors also needed to have the appropriate industry-specific auditing qualifications. This provides no evidence that might raise concerns that the auditor assignment process might bias our results (e.g., there is no evidence that auditors thought to be "more stringent" were disproportionately assigned to in-person audits). Second, including site-standard fixed effects in our model specifications holds constant the minimum auditing qualifications needed to audit a particular site for a particular standard, which mitigates the potential concern that arises in a cross-site analysis: that sites in more complex industries both require in-person audits (due to their complexity) and require audit teams with more qualifications (who are more likely to find more violations due to their additional qualifications). Third, we also learned that the audit company first decided who to assign to a particular audit, and then decided whether the audit would be conducted in-person or remotely using the process

[50] Questions in this checklist include those regarding the number of positive COVID-19 cases among audited site employees in the past weeks, whether the site has implemented social distancing guidelines in all areas, and whether there is a dedicated area for auditor's to utilize where they can maintain physical distance from others.

described above (which does not take into account who had been assigned to conduct the audit), further relieving concerns that individual auditor attributes might influence their assignment to audits or how the audit was conducted.

Results

Hypotheses 1-6. Given that the dependent variable reflects a count measure, all hypotheses are estimated using a Poisson fixed effects regression and reports standard errors clustered at the audited site. We interpret effect sizes in terms of incidence rate ratios (IRR) and average marginal effects (AME).

Table 2.3 contains the main results which test all hypotheses. Before discussing the results of the hypothesized relationships, the statistically significant effects of several control variables on audit quality warrant discussion. In Table 2.3, column 1, the positive coefficient on *staff-days (log)* indicates that audits which are assigned more staff-days to audit tend to yield more violations, possibly because the audited sites are larger (in terms of employment) and/or their management system is more complex, which presents more opportunities for auditors to uncover violations. The positive coefficient on *female on audit team* indicate that the presence of a female auditor on an audit team is associated with an improvement in audit quality, a finding consistent with Short, Toffel and Hugill (2016). Similarly, the negative coefficient on *percent outsourced* means that audit teams with a larger proportion of auditors who are not direct employees of the audit provider produce fewer violations in their audits, a result that agrees with Ibanez et al. (2022).

Table 2.3: Regression Results

	Dependent variable: Violations	(1)	(2)	(3)	(4)
H1	remote audit	-0.292**	-0.227**	-0.243**	-0.392**
		(0.026)	(0.029)	(0.027)	(0.040)
H2	remote audit X observation		-0.425**		
			(0.045)		
H3	remote audit X multi-member audit team			-0.298**	
				(0.046)	
H4	remote audit X average prior in-person site exposure (log)				0.087**
					(0.026)
	multi-member audit team	0.036	0.072	0.134**	0.031
		(0.041)	(0.047)	(0.045)	(0.041)
	average prior in-person site exposure (log)	-0.135**	-0.127**	-0.137**	-0.195**
		(0.021)	(0.023)	(0.021)	(0.028)
	focal standard advanced training	0.041	0.024	0.042	0.041
		(0.034)	(0.040)	(0.034)	(0.034)
	staff-days (log)	0.422**	0.432**	0.424**	0.423**
		(0.051)	(0.060)	(0.052)	(0.051)
	COVID time period	-0.065+	-0.241**	-0.072+	-0.068+
		(0.039)	(0.042)	(0.039)	(0.039)
	prior remote site exposure	-0.038	-0.044	-0.031	-0.051+
		(0.029)	(0.032)	(0.029)	(0.029)
	maximum auditing experience (log)	-0.033**	-0.044**	-0.033**	-0.029*
		(0.012)	(0.014)	(0.012)	(0.012)
	female on audit team	0.082*	0.075*	0.087*	0.081*
		(0.035)	(0.038)	(0.035)	(0.035)
	percent outsourced	-0.109**	-0.119**	-0.107**	-0.112**
		(0.039)	(0.044)	(0.040)	(0.040)
	multi-standard audit	-0.022	-0.023	-0.010	-0.026
		(0.041)	(0.046)	(0.042)	(0.041)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	35,247	37,631	35,247	35,247
	Number of audited sites	14,615	13,296	14,615	14,615
	Number of audited site-standards	16,986	15,388	16,986	16,986

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3 and 4, and is at the audited site-standard-detection mode for column 2. The N in model 2 is less than twice as large the N in all other models because we omit ISO 13485 audits (we cannot map violations to the primary detection mode for this standard) and because our inclusion of the three-way fixed effect (as opposed to the two-way fixed effect included in all other models) omits additional observations when there are no violations recorded within a site, standard, and detection mode. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site.
**p<0.01; *p<0.05; +p<0.10.
[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

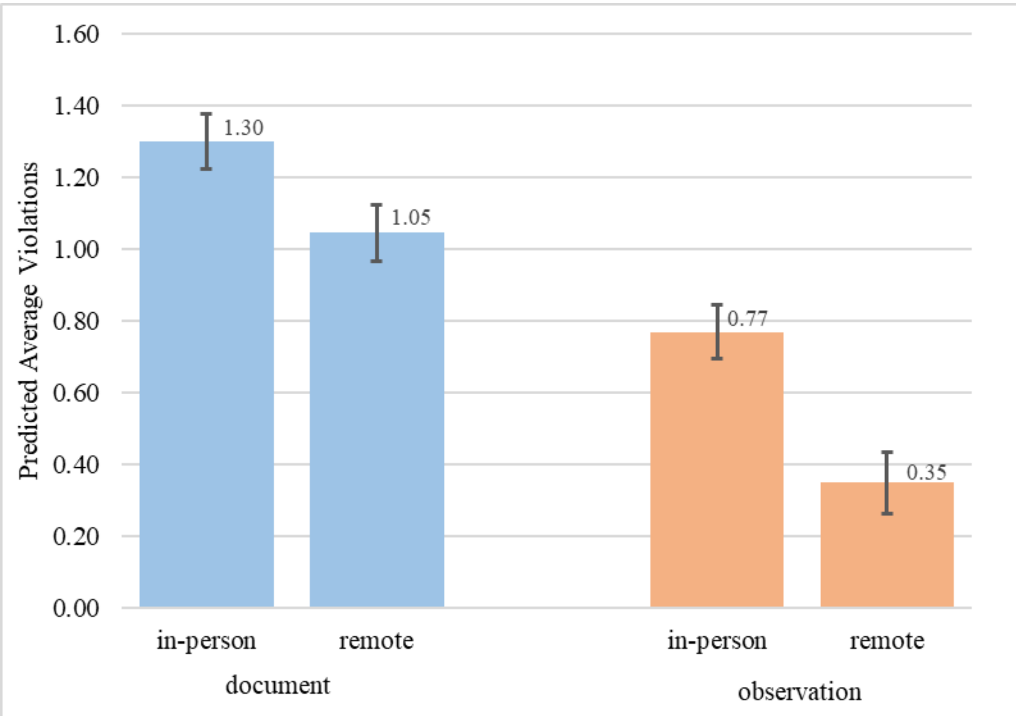
Model 1 presented in Table 2.3 tests hypothesis 1. The negative and statistically significant coefficient on *remote audit* (β = -0.292, p<0.01, IRR=0.75) indicates that remote audits yield 25% fewer violations on average than in-person audits. In terms of average marginal effects, in-person audits yield an additional 0.40 violations on average relative to remote audits, which provides support for this hypothesis.[51]

Column 2 of Table 2.3 reports our test of H2, that remote audit quality will be especially low for violations recorded primarily using an observation detection mode (vs document review). The statistically significant negative coefficient on *remote audit* (β = -0.227 $p<0.01$, IRR = 0.800) reveals that remote audits yield fewer violations than in-person audits in compliance areas primarily audited through document review. One possible explanation for this result is that these compliance areas primarily audited through document review use secondary audit methods (interviews or observation) to confirm the presence of violations, and the increased difficulty in obtaining information in remote audits through these secondary methods makes it more difficult to gather enough audit evidence to record violations. The statistically significant negative coefficient on *remote audit X observation* (β = -0.425, $p<0.01$, IRR = 0.654) supports H2, which tests one mechanism underlying H1, that a reduction in audit quality in remote audits is partially due to greater challenges in accessing information at the audited site. In support of this mechanism, this result supports that audit quality of remote audits would especially suffer among those domains where detection typically involves physical observation of the audited site. As displayed in Figure 2.1, average marginal effects indicate that, for violations primarily recorded through document review, in-person audits yield 1.30 violations on average, while remote audits yield 1.04 violations, a gap of 0.26 violations. For violations primarily recorded through audited

[51] Predicted average violations for in-person audits is 1.58 and predicted average violations for remote audits is 1.18.

site observation, the remote audit violation gap increases to 0.42 violations, as in-person audits yield 0.77 violations on average through site observation and remote audits yield 0.35 violations.

Figure 2.1: Average Predicted Effects of Remote Audit and Audit Detection Mode Interaction

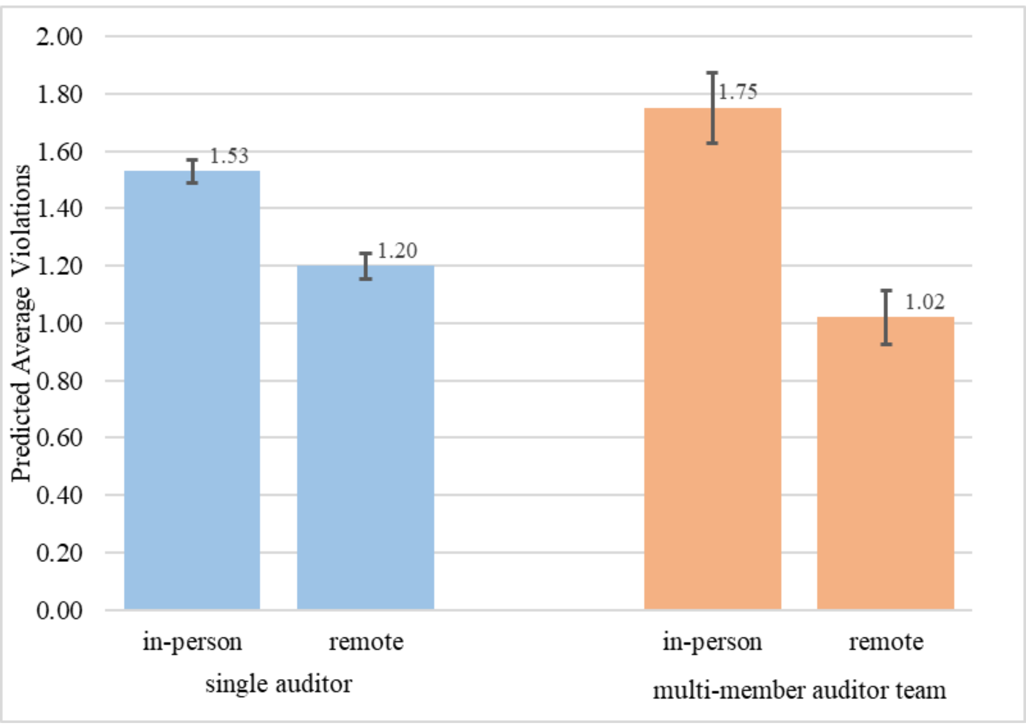


Notes: This figure relies on results from Table 2.3, column 2, and depicts predictive average effects and their 95% confidence intervals. This figure shows that the decrement of remote audits for violations identified primarily through the observation of audited site processes is significantly larger than the decrement of remote audits for violations identified primarily through document-review.

Model 3 in Table 2.3 tests hypothesis 3, which predicts that the violation gap between remote and in-person audits will be intensified for multi-auditor teams. The negative coefficient on *remote audit X multi-member team* (β = -0.298, p<0.01, IRR = 0.74) indicates that remote audits conducted by multiple-auditor teams suffer greater quality degradation than remote audits conducted by a single auditor, which supports H3. Relative to the average 22% decrease in violations reported in remote audits compared to in-person audits for single-auditor teams (*remote audit* β = -0.243, p<0.01, IRR = 0.784), this gap is significantly larger – it increases by

26% – when comparing multi-auditor audits. Average marginal effects indicate that audits with single-auditor teams see a 0.33 violation decrease when transitioning from in-person to remote audits, while this same transition leads to a 0.73 violation decrement for audits with multiple-auditor teams. (see Figure 2.2). This finding is consistent with our second proposed mechanism – that audit teams experience coordination problems in remote audits. In in-person settings, these meetings tend to occur more frequently and at more natural points throughout the audit due to the auditors' colocation. However, in remote audits, the lack of auditor colocation potentially reduces the frequency and natural timing of these meetings, which negatively impacts the quality of their work.

Figure 2.2: Average Predicted Effects of Remote Audit and Multi-Member Team Interaction

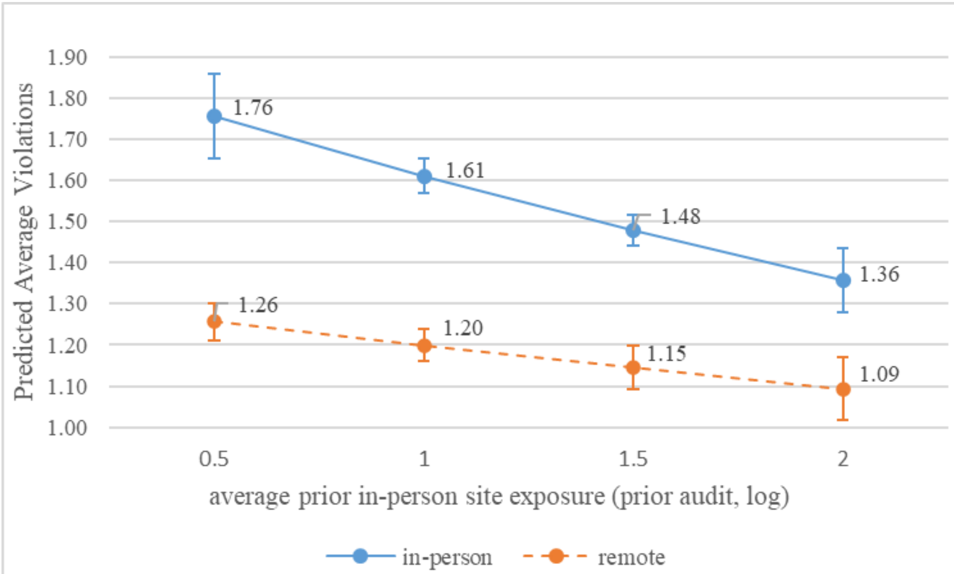


Notes: This figure relies on results from Table 2.3, column 3, and depicts predictive average effects and their 95% confidence intervals. This figure shows that the gap in predicted average violations between remote and in-person audits is significantly larger for multi-member audit teams (2+ auditor teams) than for single-auditor teams.

Column 4 in Table 2.3 reports results of our testing hypotheses 4. This tests whether the quality decrement of remote auditing compared to in-person auditing is attenuated when auditors have had more prior in-person exposure with the audited site. The negative coefficient on *average prior in-person site exposure* (β=-0.195, p<0.01, IRR = 0.823) indicates lower quality in-person audits when staffed by audit teams with more prior in-person audited site visits, a result that is consistent with prior literature (Short, Toffel, and Hugil 2016; Ball, Siemsen, and Shah 2017). The statistically significant coefficient on *remote audit X average prior in-person site exposure (*β=0.087, p<0.01, IRR = 1.091) provides evidence that an auditor having more in-person audited site visits attenuates the decrement of remote audits on audit quality, which supports H4.[52] Each one-unit increase in *average prior in-person site exposure (log)* is associated with a 17.7% reduction in in-person audit violations, but this violation gap shrinks by 9.1% in remote audits. The average marginal effects, which are displayed in Figure 2.3, show that, for example, tripling the average number of prior in-person audited site visits from approximately 2 unlogged average prior visit to 6 reduces the violation gap between remote and in-person audits by 0.14 violations on average.[53] Figure 2.3 indicates that the moderating effect of more prior exposure on audit quality is driven by a steeper decline in in-person audit quality for audit teams with more prior in-person exposure. This is unexpected, as our theory underlying H4 suggests the moderating effect would be driven by an increase in remote audit quality for teams with more prior in-person exposure.

[52] Because the maximum value of unlogged *average prior in-person site exposure* is 48, one concern is that outlier values of this variable drive our results for this variable (even after taking the natural log of this variable). To explore this concern, we estimated a model that top-codes *average prior in-person site exposure* at the 95th percentile (9 average prior in-person site visits) prior to applying the natural log transformation. As displayed in Table D-7, column 2, we find results consistent with those in Table 2.3, which suggests that our results are not driven by outlier values of *average prior in-person site exposure*.
[53] Unlogged approximations of *average prior in-person site exposure (prior audit)* are found by exponentiating the logged values (at *average prior in-person site exposure (prior audit*) = 1 and 2) and subtracting one.

Figure 2.1: Average Predicted Effects of Remote Audit and Prior Exposure Interaction



Notes: This figure relies on results from Table 2.3, column 4, and depicts predictive average effects and their 95% confidence intervals. This figure indicates that the audit quality gap is attenuated when auditors have more prior in-person site exposure. Because this variable is modeled using the natural log, the x-axis reflects logged variable value.

Robustness Tests

Our results are robust to several alternative analyses, the results of which can be found in Appendix D. First, we conduct two tests to mitigate concerns that an auditor's discretion over which specific areas of the management standard explains the negative quality decrement associated with remote audits. Management system standards contain some clauses for which auditors have discretion about whether to include them in the scope of the first or the second surveillance audit that occur between the every-three-year certification audit cycle. We refer to these as discretionary clauses and refer to those areas that must be covered in every surveillance audit as mandatory clauses. One concern that could bias our results is that auditors might choose to audit fewer discretionary clauses upon learning that the surveillance audit will be remote. A decrease in violations in remote audits could therefore be observed because auditors are choosing to audit a smaller set of discretionary clauses in remote audits.

To explore whether this concern affects our results, we pursue two approaches. First, we re-estimate our models but omit a specific sequence of surveillance audits that are at greatest risk for auditors to behave in this way. Specifically, we omit 742 site-standards (corresponding to 1,674 audits, or 5% of our sample) where the first audit in the sequence is remote and second is in-person because this sequence could enable auditors to choose fewer discretionary clauses in the first (remote) audit and then audit the larger remaining set of discretionary clauses in the second (in-person) audit. However, auditors would not exhibit this behavior for the reverse sequence, where the first audit is in-person and the second is remote, because auditors decide how to allocate discretionary clauses to surveillance audits at the time of the first surveillance audit. When the first surveillance audit is in-person, auditors will likely not anticipate the second to be remote, and so would likely allocate an equal number of discretionary clauses to both surveillance audits. Table D-1 presents these results, and we find that our results are robust to the exclusion of the audit sequence at greatest risk of containing this potential endogeneity.

Second, we re-estimate out models with the dependent variable being the number of violations that omits those violations associated with discretionary clauses. By testing our hypotheses on violations that stem from the standard set of clauses that must be audited in every surveillance audit, we avoid the possible concern that auditor discretion might be correlated with remote versus in-person audits that could bias our results. First, we omit all audits of standard ISO 13485 because we could not obtain a list of mandatory clauses (and our main results are robust to the omission of ISO 13485 audits).[54] Second, we re-estimate our models using a dependent variable (mandatory clause violations) that totals violations stemming only from those mandatory clauses. In Table D-2, we find that our results are robust to the omission of discretionary clauses.

We explore other robustness tests to mitigate concerns that decisions around variable construction produce spurious results. Because minor violations account for 97.5% of all violations in our sample, one concern is that the anomalous presence of major violations may bias our results. Therefore, we re-estimate all models using a dependent variable that includes only minor violations in Table D-3 and find that our results are robust to this dependent variable remeasurement. Our results are also robust to using average versions of the variable *maximum auditing experience (log)*. Using average versions of this variable, as seen in Table D-5, does not affect our results.

We also include an additional analysis to address potential concerns that our sample construction produced spurious results. To do so, we re-introduced the group of audits back into our sample that reflect the largest omission of audits in our sample that are unrelated to model specification decisions. Specifically, we add back 15,946 audits (corresponding to 3,552 additional site-standards) that were previously omitted from our sample because we observed bi-annual surveillance audits as opposed to the anticipated annual surveillance audit. We considered combining these multiple surveillance audits conducted in a calendar year for a single site-standard, thinking they might simply represent a surveillance audit split into parts. But our

[54] We excluded ISO 13485 audits here because we did not obtain from our data provider a list of which ISO 13485 clauses are required to be audited at every surveillance audit. In Table D-6, we report results which find that our primary models that test H1, H3, & H4 estimated on a sample that excludes ISO 13485 audits yield results very similar to our main results estimated on the full sample.

preliminary analysis led us to conclude that bundled surveillance audits are not comparable to typical once-per-year surveillance audits: the former required substantially more auditor-days than the latter, even after accounting for sites' employment level and industry, the factors that audit standards specify was determining the number of auditor-days to be allocated to each audit.

To explore whether our exclusion of these 15,946 audits potentially biased our results, we re-introduce them into our sample as 7,973 bundled audits and re-estimate all models using the increased sample. All models contained two additional variables to account for two possible influences that these additional audits could have on quality. First, we include the binary variable *bundled audit* which equals 1 if the audit reflects a bundled audit, and 0 otherwise. Second, because the two portions of the bundled audit could be performed using both remote and in-person audit formats, we include *hybrid audit* which equals 1 if the bundled audit was partially performed remotely and partially performed in-person. The results for this robustness test can be found in Table D-4, and we find that our results for all hypotheses are robust to the inclusion of these audits.

EXTENSION

Our main analysis explores whether prior in-person auditing exposure with the audited site attenuates the relationship between remote audits and quality (H4). We also explore the potential moderating relationship between another facet of auditor exposure and remote audit quality. Specifically, we questioned whether auditors that focus more of their auditing experience toward the focal audit's management standard attenuates the negative relationship between remote audits and quality. This would be consistent with well-known prior literature which documents performance benefits associated with individual workers focusing more on a specific tasks (Smith 1776; Taylor 1911; Fayol 1967). To do so, we measure *maximum management standard focus* as the percentage of staff-days conducted for the focal audit's standard (e.g., ISO 9001). Specifically, this variable counts, for each auditor, the number of prior audits performed for the focal audit's standard and divides it by the total number of prior audits performed for any standard (maximum across team). Results in Appendix Table C-1 reveal that *maximum management standard focus* has no significant association with audit quality when added into Model 1 of Table 2.3 that tests hypothesis 1. However, the positive and statistically significant coefficient on the interaction of *remote audit X maximum management standard focus* in column 2 reveals that the negative relationship between remote audits and quality is attenuated when auditors focus more of their prior auditing experience toward the focal audit's standard. This suggests that accumulating more auditing experience which is focused on a given standard can partially mitigate the information access challenges associated with remote auditing.

DISCUSSION

In this paper, we hypothesize and find that audit quality is negatively impacted when audits are conducted remotely. We theorize and find evidence of two mechanisms associated with this quality degradation. First, remote auditors' face particular challenges gathering information through direct observation, which we attribute to their vision and other senses being hampered due to technology-mediation, results in their finding and reporting fewer violations than in-person auditors. In contrast, a far smaller gap exists between remote and in-person auditors' propensity to report violations based on document-review. Second, we find that remote audits conducted by multiple auditors (versus those conducted by induvial auditors) suffer even worse quality degradation, suggesting that the collaboration necessary among multi-member audit teams to process the evidence they gather during an audit is especially challenging when conducting audits remotely. Finally, we theorize and find evidence that audit teams with more prior in-person audited site visits attenuate the negative quality decrement of remote audits.

Our findings contribute to the remote work literature by providing evidence that working remotely in ways that affect work processes (compared to working in-person) can erode the quality of the work being conducted. That several prior studies of remote work focused on settings where work processes were unaffected might explain why those studies found no difference in work quality between remote and in-person work (Bloom et al. 2015; Choudhury et al. 2021). Given that many other work settings besides auditing entail direct observation to gather critical information, such as during medical exams and other diagnostic and inspection contexts, our results should trigger special effort by those engaged in working remotely in such contexts to prevent quality loss.

We also contribute to the literature on that has examined the quality of business monitoring, which has explored how the quality of in-person auditing is affected by individual- and company-characteristics affect (e.g., Macher, Mayo, and Nickerson 2011; Duflo et al. 2013; Short, Toffel, and Hugill 2016) and inspections to assess regulatory compliance (e.g., Pierce and Toffel 2013; Ibanez and Toffel 2020). Our study builds on this work by examining how remote audits affect audit quality. Our findings are especially relevant to this literature stream given the reported rise and expected persistence remote audits and inspections in a variety of contexts, not only among management system certification audits (IAF, ILAC, and ISO 2021) but also for regulatory inspections (U.S. Department of Health and Human Services, 2022).

Our research can also aid auditing managers and regulatory inspectors who may have conflicting perceptions of remote audit quality. One recent survey of certification auditors and audited businesses found that 80% indicated that "that remote procedures [gave them] the same confidence as on-site audits," that 91% thought that organizations can benefit from remote techniques to some or a great extent and that the majority would like to see continued or increased use of remote activities in the future" (IAF, ILAC, and ISO 2021: 1-2). However, a recent report by the U.S. Department of Labor on inspections performed by Occupational Safety and Health Administration (OSHA) auditors indicated concern that "since most OSHA inspections were done remotely during the pandemic, hazards may go unidentified and unabated longer…" (U.S. Department of Labor, 2021: p 2). Our results suggest that these concerns are warranted by providing evidence that remote auditing yielded lower quality audits. Our results suggest several actions managers can take to mitigate this quality concern, including using single auditors rather than multiple auditors to conduct remote audits. By identifying auditing techniques which are more or less amenable to remote auditing, our results may also help managers design a hybrid auditing approach. Specifically, a hybrid auditing approach that allowed for remote document review, but in-person observation of the audited site, could potentially allow auditors to take advantages of some remote auditing benefits – such as increased auditing flexibility – while limiting the negative quality impacts of a fully remote audit. Moreover, building on practitioner enthusiasm to deploy hybrid auditing regimes that combine in-person and hybrid auditing approaches (IAF, ILAC, and ISO 2021), our research suggests designing hybrid auditing programs that combine in-person audits focusing on areas that require direct observation and remote audits that focus on areas assessed via document review.

Our study has limitations that highlight future research opportunities. While our sample period includes (but is not limited to) the COVID-19 pandemic and our data provider indicated that remote auditing was only used when in-person audits were infeasible, it is possible that the COVID-19 pandemic affect audited site operations in ways that could affect site compliance with management system standard requirements. For example, it is possible that sites audited remotely during the COVID-19 pandemic were operating at a lower capacity than typical operational capacity, which could naturally reduce the presence of violations because there are fewer opportunities for violation occurrence. However, to address this concern we include a *COVID-19 time period* dummy variable (that equals 1 from March-December 2020) to hold constant the effect the pandemic may have had on audit outcomes, and we have no reason to believe that operational levels varied at audited sites by audit format within the height of the COVID-19 pandemic (which would be a necessary assumption for this concern to bias our results). It is also possible that sites audited during the height of the COVID-19 pandemic were operating with unusually strong adherence to management system standards requirements, and that might explain our finding that remote audits yielded fewer violations. We have no reason to believe that, and suspect the opposite: that sites for which COVID made in-person audits infeasible were likely to be especially focused on production practices and perhaps less likely to focus on the procedural requirements of management standards making them more prone to audit violations; in such circumstances, our results should be viewed as conservative estimates of quality degradation associated with remote audits. Future work could compare remote to in-person audit quality in a different time period where the operational effects of COVID-19 are potentially less pronounced.

It is also possible that remote auditing yielded less comprehensive audits due to auditors experiencing increased mental exhaustion from doing a significantly larger portion of their work online. Mental fatigue due to increased technology use could negative impact remote audit quality given that remote audits are conducted using technology and thus would contribute toward auditors becoming even more fatigued. While our study cannot account for the potential mental costs of remote work, future research could explore the effect of technology-induced fatigue on work quality.

Incentivizing Supplier Improvement in Global Supply Chain Audits: The Role of Economic Relationships

Ashley Palmarozzo

INTRODUCTION

In many business contexts, downstream firms rely on monitoring to attest to the quality of their upstream suppliers. Popular consumer brands, such as Nike, Apple, and Walmart, monitor the ethical quality of their supply chains by assessing suppliers' compliance to specific codes of conduct. Firms are generally willing to engage in this costly form of private governance (estimated to be an $80 billion dollar industry; see AFL-CIO 2013) to safeguard company reputation (Bartley and Child 2014) and/or to ward off more stringent public regulation (Utting 2005; Barkemeyer 2009), which could be more costly than their private monitoring schemes. Perhaps for these reasons, all of *Fortune 500* companies follow some form of codes of conduct to ensure minimum quality workplace standards (McBarnet 2007).

Despite the investment in supplier monitoring, severe workplace violations persist for monitored suppliers. For example, public outcry from a large Bangladeshi factory collapse in 2013, Rana Plaza, which killed over 1,000 factory workers, sparked the formation of two initiatives to improve workplace conditions. Several years later, however, improvement in working conditions is questionable, despite interventions from brands, NGOs, and the Bangladeshi government (Saxena, S.B. 2018). This example begs the question – do private monitoring schemes improve workplace conditions? If so, when are they most effective?

Past literature diverges on whether these private monitoring programs can affect improvement in workplace conditions. One stream of literature argues that private monitoring schemes are purely ceremonial and therefore would not actually improve conditions (Boiral, 2007; Bromley and Powell 2012), arguing that monitoring programs are symbolic structures that are ultimately de-coupled from actual supplier practices (Meyer and Rowan 1977). A second stream of literature argues that suppliers can be incentivized to improve their workplace conditions using private monitoring schemes, especially when private governance is combined with strong government regulation (Rodriguez-Garavito 2005; Toffel, Short, and Ouellet 2015; Distelhorst, Hainmueller, Locke 2017) or strong institutional pressures from NGOs (Seidman 2007), labor unions (Oka 2015), or brands themselves (Oka 2010).

There are at least two components of the monitoring process that, if conducted incorrectly, could jeopardize a factory's compliance with a given code of conduct. First, the monitors themselves might be biased such that they fail to record workplace violations during factory audits. A growing stream of research on this topic finds that who pays of the audit (Duflo et al. 2013; Jiang, Stanford and Xie 2012; Kinney, Palmrose, and Scholz 2004), whether monitors have opportunities to cross-sell other products (Pierce and Toffel 2013; Causholli et al 2014), and high competition exists among auditors (Bennett et al 2013) can distort auditor incentives and produce biased audits. Studies also find that certain monitor characteristics (Short, Toffel, & Huill 2016) and monitor schedule characteristics (Ibanez and Toffel, 2020) can lead auditors to record more or fewer violations during factory audits. Second, after violations are identified and recorded, factories may be more or less willing to resolve violations. Outside of the institutional pressures described above that can enhance a supplier's willingness to comply, other studies show that certain supplier internal management structures (Bird, Short, & Toffel 2019) or supplier business practices such as lean manufacturing (Distelhorst et al. 2017) can improve supplier workplace conditions.

What is unclear in the current literature around incentivizing suppliers to close out violations is how a brand's economic relationship with suppliers influence supplier compliance incentives. Buyers could create positive improvement incentives for suppliers by rewarding them with larger purchase orders if they improve their working conditions, and this practice is encouraged by important non-governmental organizations, such as the Fair Labor Association (Fair Labor Association 2015). While there is some evidence that suppliers are willing to improve their working conditions to benefit their purchase orders from suppliers (Malesky and Mosely 2018), there is little evidence to suggest that buyers are offering such rewards to suppliers.

This paper adds to the growing literature on the buyer's role in supplier non-compliance by exploring two ways that buyers may influence labor standards compliance improvement in their suppliers. First, I explore whether a brand's role as a consumer can motivate suppliers to resolve active issues. It is possible that a brand's economic behavior as a customer can incentivize suppliers to improve their future compliance, especially given survey evidence that finds that suppliers claim to economic relationships with buyers into account when making compliance decisions (Maleskey and Mosely 2018). I therefore theorize how brands can potentially use their purchasing practices with a focal supplier – or with other competitor suppliers – to incentivize violation resolution. Specifically, I theorize that suppliers are more likely to resolve infractions when the brand represents a larger portion of their supply output because the supplier's cost of switching to an alternative buyer (supplier switching costs) is higher (if the focal buyer terminates their business relationship because of supplier noncompliance). I also theorize that suppliers will be more incentivized to resolve violations when the brand has more substitutions for them (in terms of number of alternative buyers they could use to procure the same product(s) from as the focal supplier), because the cost of switching to another supplier if the focal supplier does not comply is low (buyer switching costs).

Second, I explore whether and how brands, in their auditing capacities, are more or less effective in motivating violation resolution during supplier audits by examining whether audits conducted by auditors employed directly by the brand – as opposed to auditors employed by other auditing companies – better incentivize violation resolution. Here I theorize that brand auditors, which I refer to as second-party auditors, are more likely to persuade suppliers to resolve violations because second-party auditors are more perceived to have more authority over the supplier's economic relationship with the buyer.

Using a novel and proprietary violation-level dataset from a major apparel brand that tracks over 100,000 violations from discovery to resolution, I find that suppliers are more likely to resolve a violation in the focal audit when a larger proportion of the supplier's output went to the buyer at the time of the prior audit. I also find that the focal supplier's propensity to resolve violations is affect by the buyer's sourcing decisions with other related suppliers. Specifically, I find that suppliers are more likely to resolve a violation following an audit where the buyer has a larger number of other suppliers (in the focal supplier's industry) in their purchasing portfolio. Both of these findings are consistent with the theory that a supplier's economic relationship with a given buyer influences their future compliance decisions when supplier switching costs are high (H1) or when the buyer's switching costs are low (H2).

Against my hypothesis, I do not find evidence that auditors directly employed by the buyer (which I refer to as second-party auditors) are associated with greater odds of violation resolution in the next audit. Instead, I find that odds of violation resolution tend to be lower following second-party audits, possibly because second-party auditors tend to record more violations on average (Ibanez et al. 2023) and their focus on recording more violations leaves less time available during audits to communicate how to resolve active issues to suppliers.

This study directly contributes to the research on business partner inspections and monitoring quality. This is because most of the academic research on monitoring quality has been devoted to identifying how the process of recording violations can be biased, while few papers study the circumstances under which those recorded violations are likely to be resolved. If brands are intent on effecting improvement in supplier workplace conditions, understanding how suppliers can be incentivized to resolve active issues is as imperative as understanding how to de-bias the identification of those issues.

Literature Review

Prior literature has analyzed various factors that improve factory compliance with codes of conduct or other monitoring schemes. Institutional pressures such as strong government regulations (Rodriguez-Garavito 2005; Toffel, Short and Outllet 2015), strong labor unions (Oka 2015), a free press (Toffel, Short and Ouellet 2015), a higher density of NGOs (Seidman 2007; Short, Toffel, & Hugill 2020), or pressure from reputationally-conscious brands (Oka 2010) can galvanize factories to improve their compliance with codes of conduct. Other studies identify factory-specific internal structures that can improve factory compliance such as lean manufacturing methods (Distelhorst, Hainmueller, and Locke 2017), and the lack of piece-rate payment or other working standards certifications (Bird, Short, and Toffel 2019). One article (Short, Toffel, & Hugill 2020) suggests that certain audit types (pre-annoucned audits) and auditor types (auditors that are highly trained) increase compliance improvement rates of factories.

This study directly adds to the small but growing literature on how factories can be incentivized to improve their compliance by resolving existing violations. This study adds a layer of complexity to the incentive structure surrounding factory compliance with codes of conduct by examining how a brand's economic relationship with a given factory and how a brand's business relationship with other suppliers can "add teeth" to an otherwise voluntary program. While past studies identify institutional and supplier management factors that can induce factory improvement, I instead explore how suppliers use their economic relationship with a given buyer in their compliance decisions.

HYPOTHESES

The first two hypotheses theorize how the cost a supplier faces when switching to a different buyer (supplier switching costs) or how the perceived cost that the buyer faces when switching to a different supplier (buyer switching costs) can affect a supplier's decision to comply with their buyer's Code-of-Conduct by resolving violations in their working conditions.

Supplier Switching Costs. Prior literature has documented different market characteristics that can affect buyer or supplier behavior. Porter (2008) describes several circumstances under which buyers can incentivize suppliers to offer lower prices for exchanged goods.[55] One of these circumstances is the switching costs that suppliers would face if they lost the focal buyer, which are the costs associated with suppliers switching their supply to a different buyer. Prior literature indicates several sources of switching costs (for any entity in an economic relationship), such as when a firm in an economic relationship has larger physical or human capital investments specific to the economic relationship, or when a firm has larger search costs to find a

[55] These circumstances are bi-directional in the sense that reverse circumstances may exist where suppliers can incentivize buyers to accept higher prices for exchanged goods.

replacement firm to engage with (Klemperer 1995). Suppliers may face high switching costs because, if they lose the buyer's business, they may face costs associated with re-tooling (physical investment cost), costs of re-training employees (human capital cost), and the search costs associated with the lead time necessary to find a different buyer (Porter 2008; Chebat et al 2011).

When suppliers face high switching costs, prior literature suggests that they tend to respond by offering lower prices or improved product quality (Porter 2008). However, in the empirical context of global retail suppliers, suppliers that face high switching costs may not be able to lower prices due to already low product margins, or raise product quality because product specifications, including quality dimensions, are likely completely specified by the buyer. When suppliers face high switching costs but cannot improve on product price or quality, suppliers may respond to their increased switching costs by improving other quality dimensions, such as the quality of their working conditions. Prior survey evidence suggests that suppliers would be willing to improve their working conditions to benefit future purchase orders (Malesky and Mosely 2018), so it is possible that a supplier's switching cost factors into their future compliance decisions.[56]

The above leads to the following hypothesis:

H1: The greater the supplier's switching costs at the time of the prior audit, the higher the odds of violation resolution in the focal audit.

[56] This may lead suppliers to also be more compliant with a buyer's Code-of-Conduct in the prior audit. I address concerns that greater prior audit compliance for high supplier switching cost factories could bias their future compliance decisions in the identification section below.

Buyer Switching Costs. Suppliers may not only take their own costs of switching to a new buyer into account when making future compliance decisions, but their compliance behavior may also be influenced by the perceived cost that buyer would incur to switch to a different supplier (buyer switching costs). Prior literature suggests that search costs are an important component of buyer switching costs, because the greater the cost of finding a replacement supplier, the more inertia a buyer faces in their decision to switch to a new supplier (Liu 2006; Heide and Weiss 1995).

When a supplier perceives a buyer to have low switching costs, they may be incentivized to improve on dimensions of their supply relationship with the focal supplier. In the empirical context of global suppliers improvements on product price or product quality is infeasible, suppliers that perceive their buyers to have lower costs of switching to a different supplier may be incentivized to improve on quality dimensions other than price or product quality, such as the quality of their working conditions. Suppliers failing to improve the quality of their working conditions will likely not result in buyers defecting to other suppliers when buyers face high switching costs – that is, when few substitutable suppliers exist to produce the focal supplier's product. Conversely, when buyers have numerous other suppliers to procure the same product from, buyer switching costs are lower, and thus the likelihood of defecting to a different supplier is higher.[57]

This leads to the following hypothesis:

[57] This may lead suppliers to also be more compliant with a buyer's Code-of-Conduct in the prior audit. I address concerns that greater prior audit compliance for high buyer switching cost factories could bias their future compliance decisions in the identification section below.

H2: The lower the buyer's switching costs at the time of the supplier's prior audit, the greater the odds of violation resolution in the focal audit.

Outside of the effect that buyer or supplier switching costs may have on supplier compliance decisions, it is possible that suppliers' differing perceptions of different types of auditors may influence compliance decisions. It is possible that suppliers have different perceptions of auditors depending on the auditors' employer in two ways. First, suppliers may perceive auditors who are direct employees of the brand auditing them (called "second-party auditors") to be viewed as having greater authority over their future economic relationship with the buyer than third-party auditors who are employed by outside audit firms and who do not carry the brand name. Studies show that information is more likely to be acted upon when it comes from a perceived source of authority (Borgatti and Cross 2003; Thomas-Hunt, Ogden, and Neale 2003; Reinholt, Pederson, and Foss 2011). Therefore, suppliers may be more willing to act on information from second party auditors rather than third party auditors because they perceive them to have more sway over future purchase orders given that they carry the name of their buyer.

This leads to the following hypothesis:

H3: The odds of violation resolution will be greater following a second-party audit rather than a third-party audit.

DATA AND MEASURES

Empirical Context. I obtained factory audit data from a large global retail company, which I call TrendyStyle, that monitors factories within its supply chain to encourage factory compliance with applicable laws and compliance with the brand's ethical code of conduct.[58] TrendyStyle requires annual audits of all supplier factories and requires the same audit process and scope for all audits, regardless of the sourced product. Each audit is conducted by either its employees ("second-party auditors") or by auditors from one of several third-party audit firms ("third-party auditors"). TrendyStyle selects the auditors and pays for all audits and the vast majority of both second- and third-party auditors are local to the supplier factory's country or region.[59] Auditors find evidence for violations, or resolve existing violations, by reviewing the supplier's documents, touring the factory to observe conditions, and interviewing workers.[60] At the end of each audit, factories are assigned an overall rating (Red-Critical, Red, Amber, or Green) based on the number and severity of the violations.

Violations that an audit team discovers in the focal audit, or those that were discovered in a prior audit and are still unresolved, are documented in two places. First, the violation is described in the supplier's audit report and, second, they are again listed in an associated corrective action plan which lists the actions needed to resolve each violation.[61] In the audit's closing meeting, the auditor(s) describes to manager(s) at the supplier each violation and the corrective action the supplier should take the resolve the infraction. Once a violation has been issued, the auditors will check the status of each violation in each subsequent audit until the supplier factory that violation is resolved.

The code of conduct that the factory is audited against contains about 90 compliance elements that span 12 violation categories that track the Ethical Trading Initiative's Base Code

[58] TrendyStyle's ethical code of conduct is comparable to the Ethical Trading Initiative (ETI) categories.

[60] The number of auditor-days an audit requires is determined by the factory's size (measured by number of workers excluding management), in accordance with Sedex Members Ethical Trade Audit (SMETA) methodology.
[61] Once a violation has been recorded, the text describing the violation and the corrective action remains unchanged.

(itself based on International Labour Organization core conventions): child labor, worker knowledge of Code, worker discrimination, punitive worker treatment, environmental regulations, freedom of association, legal requirements, living wage, regular employment, working conditions, working hours, and voluntary employment.

Sample. I was provided data on over 15,000 factory audits conducted during 2007-2017 in over 50 countries.[62] The datasets contained the following information: supplier information (name, location, factory size, number of employees), audit information (auditing company, audit date, audit type, audit purpose) and information on the violations found during the audit (violation description, violation severity, violation category, and corrective action(s)).

To investigate the above hypotheses, I utilize the data at the violation-level, which means that the unit-of-analysis is at the factory-audit-violation level. Each violation will appear in the data in the audit it is discovered, then continues to appear in each subsequent audit that it remains unresolved, and then disappears in all factory audits following the one in which it is resolved.[63] The factory audits correspond to 95,344 unique factory-violations which corresponds to 185,904 factory-audit-violations (the unit of analysis).[64]

The final sample is reduced to 72,983 factory-audit-violations, and this reduction is due to a few reasons.[65] First, I omit 69,460 observations that reflect a factory's first audit in the sample because all primary independent variables use prior audit values, which I cannot observe

[62] Numerical approximations are used to help anonymize the data provider's identity.
[63] Violations may re-occur in a future factory audit, but those violations are assigned a new violation identification number and are thus treated in the data as a new violation.
[64] One factory-violation may correspond to several factory-audit-violations if the factory requires several audits to resolve it. For example, a factory that has 1 violation that is discovered in its 2nd audit and then resolved in its 4th audit will have 1 factory-audit violation, but have 3 factory-audit-violations because the violation will appear as an observation for the 2nd, 3rd, and 4th factory audit.
[65] See Table B-1 for data pipeline table.

for the first-in-sample audits of each factory. Second, I omit 43,461 factory-audit-violations that reflect a violation's first audit (the audit in which that violation is first recorded).[66] This is because suppliers cannot resolve a violation in the audit in which it is discovered, so a violation's probability of resolving a violation in its first audit is always zero.

Dependent Variable. I measure violation resolution as a binary variable *resolved*, that equals one when the violation is resolved in the focal audit, and 0 in all prior audits where the violation has been discovered but is unresolved. This measure differs slightly from prior empirical research studying factors that improve factory compliance with codes of conduct, and those studies measure improvement in a few ways. Some studies measure compliance through a factory rating (Distelhorst, Hainmueller, and Locke 2017) or by calculating a compliance rate measure (Oka 2010; Oka 2015).[67] Other studies increase the granularity of this metric by measuring improvement as the difference in labor practice scores between the prior audit and focal audit (Bird, Short, & Toffel 2019) or by measuring the percent difference in the number of violations recorded in the prior audit and focal audit (Short, Toffel, & Hugill 2020). I am able to add further granularity to this metric because, instead of using aggregate violation counts at the audit level to study changes in compliance, I can decompose compliance further to the violation unit of analysis. This additional layer of granularity allows me to more precisely control for violation characteristics that may naturally be associated with higher or lower odds of violation resolution, such as a violation's severity and functional category.

---

[66] For example, if in a factory's second audit, there are 5 violations that were discovered in the prior audit and that are unresolved, and auditors discover one new violation in the factory's 2nd audit, I would omit the second audit observation for the newly discovered violation, but include the second audit observations for the 5 existing violations.

[67] This is found by identifying and adding up all code of conduct categories that the factory was compliant for and dividing by the total number of code of conduct categories.

Of the 48,279 unique factory-violations recorded in the final sample, 32,770 are eventually resolved, about 68%. 84% of the violations are classified as major violations, with the remaining 16% classified as minor.[68] On average violations are resolved in the second audit after which they were discovered, and this difference is similar for major and minor violations.[69]

Primary Independent Variables. As described above, all hypotheses use prior audit values because each hypothesis theorizes how various conditions will induce suppliers to resolve active issues following an audit. To test hypotheses 1 and 2, I develop two proxies of supplier (H1) or buyer (H2) switching costs. Prior literature indicates several sources of switching costs for any firm in an economic relationship with another firm, such as a firm in such an economic relationship having larger physical investments, or when firms face greater search costs (Klemperer 1995). However, many note a significant gap between theoretical switching cost scholarship and empirical switching cost scholarship (Gomez and Maicas 2011; Grzybowski 2008; Chen and Hitt 2007). Therefore, prior empirical switching cost literature has not converged on common measurements of switching cost constructs. Instead, scholars tend to do one of two things. First, some scholars use an event, such as the enactment of a policy, to model a universal increase (or decrease) in switching costs for some economic agent (e.g., customers) (Abolfathi, Santamaria, and Williams 2021). Second, some scholars focus on measuring one (or more) sub-category of switching costs that are especially relevant to their empirical setting (eg., Brush, Dango, and O'Brien 2012; Carnahan and Somaya 2013). For example, Brush, Dangol and O'Brien (2012) assumes the cost that customers face when switching to a new bank (with online banking features) is primarily a function of their learning costs associated with the focal bank

[68] The classification of violation severity relies on TrendyStyle's violation severity classification. Approximately 66% of major violations are eventually resolved, while 76% of minor violations are eventually resolved.
[69] See Figure G-2 for distribution of resolved violations by time to resolution.

(and they measure focal bank learning costs as the number of years a customer has had an account at the focal bank). I utilize the latter method used by scholars by considering the relevant sub-category of switching costs that I consider to have a primary impact on switching costs in my empirical setting.

In my empirical context, the cost that suppliers face when switching to a different buyer is likely driven by the portion of their output going toward the focal buyer because suppliers that concentrate more of their output toward a single buyer have larger specific physical investments (e.g., machinery) and larger human capital investments (e.g., training workers) with that buyer. Additionally, search costs are likely higher for these suppliers, as they must fill a larger volume of capacity if they lose the focal buyer. Therefore, to test hypothesis 1, the supplier's switching costs are measured by *percent supplied (prior audit)* that equals the factory's percentage output that is supplied to the brand as recorded in the last audit.[70]

I also assume, given my empirical context, that the buyer's search costs (the cost incurred to find a different supplier to procure from) are the primary drivers of buyer switching costs in this empirical context. This is because I assume that the buyer does not own any relationship-specific assets, nor are their products unique in ways that could lead to increases in switching costs due to buyer lock-in, so physical or human capital investments do not primarily explain the buyer's switching costs. I proxy the buyer's search costs by measuring the number of alternative suppliers that the buyer could use to replace the focal supplier, as the more alternative supply options a firm has, the lower the search cost to replace the existing supplier. Specifically, I measure *number of alternative suppliers (prior audit)* as the number of other suppliers in the

[70] This variable is recoded to 0 in 4,859 observations where this variable is missing, and I model a binary variable which equals 1 when missing (0 otherwise). Omitting missing observations does not materially affect results.

focal supplier's industry that TrendyStyle audits within the year that precedes the prior audit, logged to reduce skew.[71] While the supplier's industry is not explicit in the dataset, I am able to impute which suppliers the buyer could switch their focal supplier's order to by utilizing a field that contains product code numbers that the brand uses in product procurement. I consider any two suppliers with at least one identical product code (of which there are at maximum 24) to be in the same industry, regardless of their geographic proximity. I use the product code field to impute industry for two reasons. First, conversations with the brand describe this field as a reasonable proxy for supplier industry. Second, suppliers should consider themselves competitors of one another if they produce the same category of product for the brand. Discussion of alternative measures can be found in the robustness section of this paper.

To test hypothesis 3, second-party, or in-house audits, are identified by *second-party (prior audit)*, that equals 1 when the lagged audit is performed by second-party auditors, and 0 when the audit was performed by outsourced third-party auditors..

Control variables. I include several control variables to capture audit-, violation-, and factory-level characteristics which may be correlated with the primary independent variables and the outcome measure. I include two binary variables that equal 1 if the prior or focal audit was unannounced, called *unannounced (prior audit)* and *unannounced*. I construct a variable that equals the factory's prior audit social audit rating (*social audit rating (prior audit)).* This variable is equal to one of 4 categorical ratings assigned to the factory at the end of each audit – Red-

[71] This approach assumes that the number of suppliers in the focal supplier's industry that TrendyStyle audited is a reasonable proxy for the overall number of suppliers in that industry that TrendyStyle might consider procuring from (including those it didn't audit over the past year). This assumption appears reasonable because, all else equal, TrendyStyle faces lower transaction costs (by facing lower costs associated with auditing) when switching procurement to a supplier that is already up-to-date on auditing requirements rather than switching procurement to a supplier that will require additional auditing (either because TrendyStyle has never worked with the supplier before or because their most recent audit took place over a year ago and so is not up-to-date).

Critical, Red, Amber, and Green – and reflects a factory's overall compliance with its Code of Conduct. I measure the number of violations in the prior audit using *number of recorded violations (prior audit)*. I construct three binary variables to reflect the prior audit's scope: *new factory audit (prior audit)* (the first audit a factory receives), *re-audit (prior audit)* (each subsequent audit after the first), and *follow-up audit (prior audit)* (the type of audit performed after "red-critical" or "red" rated audits whose focus is to close out the more serious violations).[72] I include focal audit versions of *re-audit* and *follow-up audit* in the model.[73]

Using the provided first names of lead auditors and the country in which they are auditing, I use an Application Programming Interface (API) called Genderize.io to predict the lead auditor gender in both the prior and focal audit.[74,] Called *female lead auditor (prior audit)* and *female lead auditor*, values range from 0 (male lead auditor) to 1 (female lead auditor).[75] I also measure prior and focal audit versions of *first time lead auditor*, that equals 1 if the lead auditor has never visited the factory before.

I measure factory size as the number of factory workers recorded in the prior audit (called *total workers (prior audit)*), which I log to reduce skew.[76] I measure *percent female workers (prior audit)* as the percentage of workers in the factory that are female at the time of the prior audit.[77] I measure *factory tenure (prior audit)* as the number of years that the factory supplied to

[72] These variables are recoded to 0 for 9,643 observations where the prior audit's scope is missing. Because these categories are mutually exclusive, I omit *re-audit (prior audit)* from the regression specification.
[73] These variables are recoded to 0 for 12,677 observations where the focal audit's scope is missing. Because these categories are mutually exclusive, I omit *re-audit* from the regression specification.
[74] The API uses profile information across major social networks to predict gender based on first name. More information can be found here: https://genderize.io.
[75] This variable is recoded to 0 for 10,238 observations where the lead auditor name is missing and a separate binary variable is constructed that equals 1 when missing.
[76] This variable is recoded to 0 for 3,688 observations where factory size information is missing and a separate binary variable is constructed that equals 1 when missing.
[77] This variable is recoded to 0 for 3,899 observations where worker gender is missing and a separate binary variable is constructed that equals 1 when missing.

the brand at the time of the prior audit.[78] Because factory size, worker composition, and tenure are measures that tend to be stable over time, I do not include focal versions of these variables in the model to avoid multicollinearity issues.

I constructed *violation vintage* which counts the number of audits that the violation has been active for at the time of the focal audit.[79] I constructed *audit sequence* as a count variable which counts the how many times the factory has been audited at the time of the focal audit (1^{st} audit, 2^{nd} audit, etc). I constructed a series of dummy variables to reflect a supplier's industry. These variables are constructed by creating a binary variable for each of the 54 product codes a given supplier could be associated with. Each of these product code binary variables equals 1 if the supplier is linked to that product code, and 0 otherwise.

Table 3.1 displays the summary statistics for the data and Table 3.2 displays the correlation matrix for the above variables.

[78] This variable is recoded to 0 for 7,350 observations where the factory's supply start year is missing and a separate binary variable is constructed that equals 1 when missing. This variable is constructed by taking the difference between the year of the prior audit and the year provided in the data as the factory's first supply year, which can be a year prior to the start year of my data sample. This measure assumes that the factory continuously supplies to TrendyStyle after the supply start year listed in the data, and thus cannot incorporate instances where a supplier stop and re-start their supply to TrendyStyle. This variable may therefore overestimate the number of years that a factory has supplied to TrendyStyle.

[79] For example, if a violation was opened in a factory's 2^{st} audit and is closed in their 4^{th} audit, then at the time of the 4^{th} audit, *violation vintage* equals 3.

Table 3.1: Summary Statistics

Variable	mean	sd	min	max	count
resolved	0.45	0.50	0	1	72983
percent supplied (prior audit)[1]	0.23	0.28	0	1	68124
number of alternative suppliers (prior audit)	1126.59	654.19	0	1909	72983
number of alternative suppliers (prior audit)[2]	5.91	2.70	0	7.55	72983
second-party (prior audit)	0.35	0.48	0	1	72983
second-party	0.26	0.44	0	1	72983
unannounced (prior audit)	0.19	0.39	0	1	72983
unannounced	0.28	0.45	0	1	72983
number of recorded violations (prior audit)	13.86	9.93	1	76	72983
new factory audit (prior audit) [1]	0.19	0.39	0	1	63340
re-audit (prior audit) [1]	0.51	0.50	0	1	63340
follow-up audit (prior audit) [1]	0.30	0.46	0	1	63340
re-audit[1]	0.52	0.50	0	1	60306
follow-up audit[1]	0.48	0.50	0	1	60306
female lead auditor (prior audit) [1]	0.46	0.47	0	1	62745
female lead auditor[1]	0.40	0.45	0	1	62745
first time lead auditor (prior audit)	0.83	0.37	0	1	72983
first time lead auditor	0.77	0.42	0	1	72983
total workers (prior audit) [1,2]	5.60	1.29	0	9.25	69295
percent female workers (prior audit) [1]	0.55	0.23	0	1	69084
factory tenure (prior audit) [1]	3.16	3.62	0	25	65633
violation vintage	2.65	1.04	2	6	72983
audit sequence	2014.53	1.90	2011	2017	72983
audit year	4.35	1.99	2	8	72983
Major violation	0.85	0.36	0	1	72983

Notes: Sample reflects 72,983 factory-audit-violation observations of 48,279 unique violations in 2,626 unique factories. [1] When missing, this variable is set to 0 and a separate binary variable that equals 1 when missing is modeled. [2] reflects variables that use the natural log.

Table 3.2: Correlation Matrix

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
resolved	1	1.00																				
percent supplied (prior audit)[1]	2	0.06	1.00																			
number of alternative suppliers (prior audit)[2]	3	-0.02	0.11	1.00																		
second-party (prior audit)	4	-0.01	-0.07	-0.12	1.00																	
second-party	5	0.03	0.02	-0.07	0.22	1.00																
unannounced (prior audit)	6	0.00	0.10	0.11	0.12	0.09	1.00															
unannounced	7	-0.04	0.05	0.09	0.13	0.21	0.41	1.00														
number of recorded violations (prior audit)	8	0.01	-0.02	-0.08	0.07	0.13	0.15	0.05	1.00													
new factory audit (prior audit) [1]	9	-0.06	-0.29	-0.10	0.26	-0.11	-0.12	0.01	-0.14	1.00												
re-audit (prior audit) [1]	10	-0.03	0.13	0.22	-0.27	-0.05	0.07	0.09	0.00	-0.40	1.00											
follow-up audit (prior audit) [1]	11	0.02	0.15	0.10	-0.03	0.04	0.16	0.06	0.04	-0.26	-0.53	1.00										
re-audit[1]	12	-0.12	-0.09	0.15	0.00	-0.22	-0.14	-0.04	-0.29	0.35	-0.11	-0.03	1.00									
follow-up audit[1]	13	0.09	0.14	0.04	-0.09	0.12	0.20	0.12	0.22	-0.29	0.24	0.11	-0.87	1.00								
female lead auditor (prior audit) [1]	14	0.00	0.01	0.06	-0.18	-0.13	-0.08	-0.07	0.03	-0.03	0.07	-0.01	0.02	0.03	1.00							
female lead auditor[1]	15	-0.02	0.00	0.05	-0.15	-0.19	-0.09	-0.09	-0.01	0.02	0.04	-0.01	0.08	-0.03	0.24	1.00						
first time lead auditor (prior audit)	16	0.02	-0.02	-0.08	-0.12	-0.04	-0.13	-0.10	-0.07	0.08	0.02	-0.12	0.00	-0.01	0.04	0.03	1.00					
first time lead auditor	17	-0.03	0.00	0.04	-0.04	-0.23	-0.11	-0.06	-0.20	0.12	0.00	-0.03	0.11	-0.04	0.04	0.03	0.16	1.00				
total workers (prior audit) [1,2]	18	0.02	0.05	0.06	0.02	0.02	0.18	0.14	0.06	-0.03	0.03	0.12	-0.03	0.13	-0.08	-0.05	-0.06	-0.01	1.00			
percent female workers (prior audit) [1]	19	-0.09	0.04	0.05	-0.12	-0.16	-0.10	-0.09	-0.05	0.07	0.09	-0.05	0.13	-0.05	0.10	0.12	0.11	0.15	0.19	1.00		
factory tenure (prior audit) [1]	20	0.05	0.21	0.12	-0.09	0.03	0.05	0.04	-0.04	-0.32	0.11	0.19	-0.11	0.14	-0.01	-0.03	-0.10	-0.06	0.10	-0.07	1.00	
Major violation	21	-0.08	0.00	0.00	-0.01	0.02	0.02	0.00	0.08	-0.04	0.02	0.04	-0.06	0.06	-0.02	-0.03	0.02	0.03	-0.02	0.02	0.02	1.00
Social audit rating (prior audit)	22	-0.09	-0.09	0.05	0.01	-0.27	-0.20	-0.17	-0.34	0.32	-0.10	-0.15	0.65	-0.61	0.01	0.81	0.01	0.10	-0.12	0.06	-0.10	-0.08

Notes: Sample reflects 72,983 factory-audit-violation observations of 48,279 unique violations in 2,626 unique factories. [1] When missing, this variable is set to 0 and a separate binary variable that equals 1 when missing is modeled. [2] reflects variables that use the natural log.

EMPIRICAL MODEL AND RESULTS

Empirical Specification. I estimate the effect of the hypothesized variables using a multilevel mixed-effects logistic regression with random effects at the factory and audit (clustering standard errors at the factory) to test my hypotheses. I utilize this model for a few reasons. First, I use a logistic regression because of the binary format of the dependent variable, *resolved*.[80] Second, I include two levels of random effects--audits nested within factories--to account for the correlation of factory-audit-violation observations within the set of observations pertaining to each factory audit and within the broader set of observations pertaining to each factory. Including random effects assumes that factory-audit-level and factory-level unobserved heterogeneity is uncorrelated with the independent variables (Greene, 2008). However, I include random effects as opposed to including factory fixed effects because factory fixed effects would result in right censoring of observations, where factories that do not resolve at least one violation by the end of the panel would be omitted from the estimation sample, and this could bias results.[81] Third, I cluster standard errors by factory to account for potential correlation among each factory's observations.

I therefore estimate the below model:[82]

$$logit\big(\Pr[Y_{i,v,t} = 1]\big) = \beta_0 + \beta_1 X_{i,t-1} + \beta_2 \lambda_{\mathrm{i,t-1}} + \beta_3 \delta_{i,t} + \beta_4 \phi_{i,v,t} + \alpha_i + \gamma_{i,t} + \mu_t + \pi_i + \theta_i + \Omega_{i,v,t} + \epsilon_{i,v,t}$$

Where $Y_{i,v,t}$ represents my dependent variable, *resolved*, which equals 1 when violation *v* in audit *t* for factory *i* is resolved, and 0 otherwise. $X_{i,t-1}$ represents the lagged primary independent

[80] This is the model choice often used in discrete time hazard scenarios. *See* Allison (2014).
[81] Including factory fixed effects omits an additional 437 audits (a 5% decrease from the 8,616 audits included in the mixed effects model) from 344 factories (a 13% decrease from 2,626 factories included in the mixed effects model.
[82] I use the *melogit()* command in Stata.

variables: *percent supplied (prior audit), number of alternative suppliers (prior audit),* and *in-house (prior audit)*. $\lambda_{i,t-1}$ represents a vector of the audit-level lagged control variables, $\delta_{i,t}$ represents a vector of the audit-level focal control variables, and $\phi_{i,v,t}$ represents violation-level control variables that are modeled as fixed effects. α_i and $\gamma_{i,t}$ represents factory and audit random effects, respectively. π_i represent supplier industry fixed effects, Θ_i represents supplier country fixed effects, μ_t represents audit year fixed effects, and $\Omega_{i,v,t}$ represent violation vintage effects.

$\lambda_{i,t-1}$ and $\delta_{i,t}$ include several control variables. I include *unannounced (prior audit)* because past research shows that unannounced audits are associated with lower rates of improvement (Short, Toffel, and Hugill 2020). I also include focal versions of this variable, called *unannounced*, because prior research finds that auditors tend to record more violations in unannounced audits (Short, Toffel, & Hugill, 2016). I include two binary variable to control for the prior audit's scope – *new factory audit (prior audit)* and *follow-up audit (prior audit)* – to account for the possibility that the auditors that conduct prior audits of differing scope could place different levels of emphasis on recording more violations vs resolving existing violations.[83] I also include one focal audit variable that reflects audit scope – called *follow-up audit*– for similar reasons.[84] I also include the prior audit's social audit rating (*social audit rating (prior audit)*) to account for the possibility that factories that appear to be less compliant (in terms of receiving a worse social audit rating) may also be less likely to comply with the brand's Code of Conduct in the future.

At the audit team level, I include prior and focal audit lead auditor gender – called *female lead auditor (prior audit)* and *female lead auditor* - because prior research suggests that female

[83] The omitted category is *re-audit (prior audit)*
[84] My decision to omit a factory's first audit from the sample means that focal new factory audits are also omitted from the sample. The omitted focal audit scope category is *re-audit*.

auditors tend to audit with greater levels of scrutiny and thus record more violations (Short, Toffel, and Hugill 2016). The gender makeup of the audit team could also influence violation resolution because female auditors may also apply greater levels of scrutiny when examining whether a violation is fully resolved. I include prior and focal audit versions of *first time lead auditor* because auditors that have not visited the factory previously tend to find more violations in factory audits (Short, Toffel, and Hugill 2016).

Prior audit factory size, called *total workers (prior audit),* is included to account for the effect that a factory's size may have on their ability to resolve violations. I include *number of recorded violations (prior audit)* to account for the possibility that a supplier that has more violations in the last audit has more opportunity to resolve violations, and likely faces greater costs in resolving all active issues, which may influence their likelihood to resolve any given violation in the next audit. I also include *audit sequence* because factories that have been audited more times may be more likely to resolve violations because they have had more prior audits to accumulate knowledge on how to resolve violations.

$\phi_{i,v,t}$ includes two variables that are modeled as fixed effects. First, I include separate fixed effects for the violation's severity (major or minor). Second, I include the violation's sub-category within the data provider's Code of Conduct (93 categories) which reflects the most detailed functional category that each violation is assigned to.[85] I do so to account for the fact that violations may naturally vary in their resolution difficulty (for example, major violations may take a larger investment of time and financial resources to resolve).

[85] $\phi_{i,v,t}$ includes 93 fixed effects: 1 fixed effect for the violation's severity ("minor" is the omitted category), and 92 fixed effects for the violation's functional sub-category ("working conditions, fire safety" is omitted category).

I also include a violation vintage fixed effect (using $\Omega_{i,v,t}$), which equals the violation's nth audit (e.g., a violation's 1^{st}, 2^{nd}, or 3^{rd} audit), to account for the possibility that violations that have been open for longer may be more difficult to resolve. I include audit year fixed effects (using μ_t) to account for any secular trends that may impact a factory's propensity to resolve violations. Finally, I include supplier country fixed effects (using Θ_i) and supplier industry fixed effects (using π_i) to account for possible differences at the country or industry which could affect resolution rates.

Identification Strategy. Hypotheses 1 and 2 theorize that suppliers that face greater supplier switching costs (H1) or that perceive TrendyStyle to have lower buyer switching costs (H2) are more likely to resolve violations in the next audit. A primary assumption made for hypotheses 1 and 2 is that the mechanisms theorized to affect a factory's resolution propensity of a given violation from the prior to focal audit does not also affect a factory's compliance in the prior in ways that could bias my analysis. It is possible that prior audit switching costs do not just motivate suppliers to resolve violations in the next audit, but they may also prompt suppliers to be more compliant with the brand's Code of Conduct at the time of the prior audit. This could be a concern if the set of violations in the prior audit are significantly different for factories that fall on either end of the supplier (or buyer) switching cost distribution. Below I separately examine, for each hypothesis, violation characteristics of factories on either end the switching cost distribution to explore whether these differences could bias my results.

Conceptually, it is possible that if this bias is present in the data, it may have an attenuating effect on violation resolution propensity. Lower supplier switching cost (or higher buyer switching cost) factories should have fewer violations at the end of each audit, and the violations that persist over time could be ones that are more difficult to resolve. This bias would then attenuate the results of both H1 and H2 because there I argue that lower supplier switching cost (higher buyer switching cost) factories, in terms of their prior audit switching costs, should be more likely to resolve violations. The opposite scenario is less likely, that the violations that are identified at more compliant factories are easier to resolve, because lower supplier switching cost (higher buyer switching cost) factories should follow TrendyStyle's Code of Conduct more closely and should therefore be more likely to immediately resolve those "easy" violations if they arise between audits.

To empirically explore whether the set of violations in the prior audit are significantly different for factories that fall on either end of the supplier (or buyer) switching cost distribution (which could bias my results), I first examine whether high supplier switching cost factories (low buyer switching cost factories) are more compliant in the prior audit by comparing whether the number of prior audit violations significantly differs across low versus high switching costs. I then examine two dimensions of prior audit violation characteristics across low vs high supplier switching costs (or buyer switching costs): violation severity and functional category. I examine these two dimensions because differences within either dimension across the switching cost distribution will affect both the types of violations present at the factory at the time of the prior audit and may affect violation resolution odds.[86]

For supplier switching costs, I consider a factory as facing higher switching costs if they fall above the median value of *percent supplied (prior audit)*, which is 20%. As displayed in Appendix Table F-1, I find that the distribution of prior audit violations, *recorded violations (prior audit)*, is significantly different across low versus high suppliers switching costs, as

[86] It is possible that differences in other prior audit variables, such as auditor characteristics, could affect the set of violations present at factories. However, the effect of audit prior audit variables on the set of prior audit violations should be reflected in the prior audit violation characteristics I can observe and compare.

measured by a chi2 test of independence (chi2(55) = 78.84, p-value= 0.019).[87] In Table F-2, which presents the mean prior audit violations by switching cost category, I find that audits of factories with higher switching costs have 0.2 more prior audit violations on average, a difference that is statistically significant according to a two-tailed t-test (p-value= 0.0175). This suggests that factories with higher supplier switching costs are more compliant in the prior audit.

Turning to a comparison of violation characteristics, I find in Table F-2 that major violations from high supplier switching cost factories occur with 84% frequency versus 83% frequency at low supplier switching cost factories, and a two-tailed proportional t-test indicates that this difference is statistically significant with $p<0.05$. In Table F-3, I do not find that the distribution of violations by functional category is significantly different for low versus high switching cost categories as the p-value associated with a chi-square test of independence reveals $p>0.05$ (chi2(11) = 12.07, pr = 0.359). This investigation provides some evidence that factories with higher supplier switching costs are more compliant at the time of the prior audit in terms of their violation count. However, because the observed violation characteristics do not significantly differ (in terms of their severity or functional category distribution) across suppliers with low and high supplier switching costs, I do not find evidence that factories with higher supplier switching costs have significantly different violation sets in the prior audit that could bias their violation resolution propensity, by inflating their odds of violation resolution, in the focal audit. To the extent that major violations are more difficult to resolve than minor violations, this analysis implies that high switching cost factories may have marginally greater difficulty resolving violations, which would attenuate the hypothesized effect in H1.

[87] The unit of analysis for this test is the factory audit, because an audit's violation count is determined at this level.

For buyer switching costs, I consider a factory as facing lower buyer switching costs (which could induce greater prior audit compliance) if they fall above the median value of *number of alternative suppliers (prior) audit*, which, unlogged, is about 1,200 suppliers. In Table F-4, I find that the distribution of prior audit violations, *recorded violations (prior audit)*, is significantly different across low versus high buyer switching costs, as measured by a chi2 test of independence (chi2(55) = 80.24, p-value= 0.015). In Table F-5, audits of factories with lower buyer switching costs have 1.3 fewer prior audit violations on average, a difference that is statistically significant according to a two-tailed t-test (p-value= 0.000). These tests together suggest that factories with lower buyer switching costs are more compliant in the prior audit.

Examining differences in violation characteristics, in Table F-6, I do not find that the distribution of violations across functional categories significantly differs for violations located in factories that have high versus low buyer switching costs (chi2(11) = 12.348, p-value = 0.34). Also, as recorded in Table F-5, I do not find evidence that major violations occur with more or less frequency across low and high buyer switching costs (according to a two-tailed t-test). Therefore, I do not find significant evidence that the violations that result from factories that face different levels of supplier switching costs or buyer switching costs differ in ways that could bias the results for hypotheses 1 and 2.

When testing hypothesis 1 (*percent supplied (prior audit)*), a second assumption made is that TrendyStyle's procurement decisions at a given factory are plausibly unrelated to that factory's average compliance levels. If TrendyStyle decided to place larger future orders (in terms of volume) from factories with higher focal compliance, this unobserved procurement strategy could positively bias the coefficient on *percent supplied (prior audit)*. Interviews with TrendyStyle auditing managers reveal that the procurement division and the auditing division of the firm are typically siloed from one another, which makes the likelihood of this potential confounder biasing results low.[88] However, to explore this concern empirically, I employ two tests.[89] First, as displayed in Table F-7, I compare the lag audit violation distribution by focal audit factories with low supplier switching costs versus high supplier switching costs (about 15%), and a chi2 test of independence indicates no significant difference in prior audit violation distribution across high versus low supplier switching cost (chi2(55) = 64.25, p-value = 0.18). Second, in Table F-8, I estimate a linear model that regresses focal audit *percent supplied* on *number of recorded violations (prior audit)* and includes all control variables from the main model (with factory fixed effects). The near-zero statistically insignificant coefficient ($\beta = 0.001$, $p>0.10$) on *number of recorded violations (prior audit)*, along with the non-significant chi2 test statistic, supports the validity of the identification assumption that prior audit compliance is plausibly unrelated to future procurement levels.

For hypothesis 3 (*second-party (prior audit)*) to yield unbiased results, the assignment of second- versus third-party auditors to factory audits should not be correlated with any unobserved factory characteristics that could affect such assignment and violation resolution propensity. Prior analysis reveals that second-party auditors may be more likely to be assigned to audit factories with more prior violations and worse prior social audit ratings, but I observe both of those dimensions and include both of those variables as control variables in the model. It's also unclear which direction the effect would be biased – worse performing factories may be more resistant to change, but they may also have the most opportunity space for improvement.

[88] One potential interaction between these divisions is that the procurement division is prohibited from placing a future order at any factory that receives a "Red-Critical" rating until that factory resolves the egregious violation(s) that led to that rating. Roughly 5.7% of factory audits result in a "Red-Critical" rating in this sample.
[89] The unit of analysis in both tests is the factory audit because that is the level at which both prior audit violations and supply percentage is determined.

Additionally, it's possible that different auditor parties record different types of violations that vary in how difficult they are to resolve. This could positively bias the hypothesized effect I seek to test if third-party auditors tend to find more violations that are more difficult to resolve. While I have no reason to suggest this identification threat is true, I include violation severity and Code of Conduct functional sub-category fixed effects in my model specification, and both of these sets of fixed effects should reasonably control for variation in the average difficulty to resolve specific violations.

Results. Table 3.3 presents the main results, which uses a multilevel mixed-effects logistic regression with robust standard errors clustered at the factory to test all hypotheses. All regressions include fixed effects for violation severity, violation functional sub-category, and audit year. Column 1 of Table 3.3 presents results of a model that estimates the effect of the control variables on violation resolution propensity without the inclusion of the hypothesized variables. Column 2 present results from re-estimating the model including the hypothesized variables, and this inclusion does not materially affect the coefficients on almost all control variable.[90] The incremental addition of country fixed effects (column 3) and industry fixed effects (column 4) affects the statistical significance of coefficients on many control variables and on the third hypothesized variable, but does not materially affect the results from the first two hypothesized variables. Columns 5 and 6 provide the odds ratio (column 5) and average marginal effects (column 6) associated with the results in column 4.

[90] The coefficient on *new factory audit (prior audit)* loses statistical significance with the inclusion of the hypothesized variables, and this is due to the inclusion of *second-party (prior audit),* as the data provider tends to send second-party auditors to perform a new factory's first audit.

Table 3.3: Regression Results

	DEPENDENT VARIABLE = resolved	(1)	(2)	(3)	(4)	OR	AME
H1	percent supplied (prior audit)[1]		0.393***	0.171**	0.200***	1.22	0.03
			(0.071)	(0.069)	(0.072)		
H2	number of alternative suppliers (prior audit)[2]		0.033***	0.024***	0.022***	1.02	0.00
			(0.008)	(0.008)	(0.008)		
H3	second-party (prior audit)		-0.036	-0.089**	-0.097**	0.91	-0.02
			(0.044)	(0.043)	(0.043)		
	second-party	-0.075	-0.069	-0.134***	-0.143***	0.87	-0.02
		(0.046)	(0.046)	(0.046)	(0.046)		
	unannounced (prior audit)	0.027	0.021	0.032	0.028	1.03	0.01
		(0.054)	(0.054)	(0.054)	(0.054)		
	unannounced	-0.148***	-0.142***	-0.234***	-0.234***	0.79	-0.04
		(0.052)	(0.052)	(0.053)	(0.053)		
	number of recorded violations (prior audit)	-0.023***	-0.022***	-0.014***	-0.013***	0.99	-0.00
		(0.003)	(0.003)	(0.003)	(0.003)		
	new factory audit (prior audit) [1]	-0.130**	-0.017	-0.054	-0.044	0.96	-0.01
		(0.057)	(0.062)	(0.061)	(0.061)		
	follow-up audit (prior audit) [1]	0.116***	0.113**	0.052	0.054	1.06	0.01
		(0.044)	(0.044)	(0.044)	(0.044)		
	follow-up audit[1]	0.161***	0.159***	0.063	0.055	1.06	0.01
		(0.058)	(0.058)	(0.057)	(0.057)		
	female lead auditor (prior audit) [1]	0.037	0.029	0.023	0.022	1.02	0.00
		(0.039)	(0.038)	(0.038)	(0.038)		
	female lead auditor[1]	-0.087**	-0.089**	-0.020	-0.026	0.97	-0.00
		(0.043)	(0.042)	(0.042)	(0.042)		
	first time lead auditor (prior audit)	0.023	0.020	0.111**	0.108**	1.11	0.02
		(0.051)	(0.052)	(0.052)	(0.052)		
	first time lead auditor	-0.026	-0.024	0.076	0.074	1.08	0.01
		(0.047)	(0.047)	(0.049)	(0.049)		
	total workers log (prior audit) [1,2]	0.002	0.006	0.046***	0.048***	1.05	0.01
		(0.014)	(0.013)	(0.015)	(0.015)		
	percent female workers (prior audit) [1]	-0.426***	-0.439***	0.150	0.133	1.14	0.02
		(0.122)	(0.121)	(0.127)	(0.128)		
	factory tenure (prior audit) [1]	0.002	-0.001	-0.005	-0.007	0.99	-0.00
		(0.007)	(0.007)	(0.006)	(0.006)		
	Major violation	-0.210***	-0.208***	-0.177***	-0.177***	0.84	-0.03
		(0.031)	(0.031)	(0.031)	(0.031)		
	audit sequence	0.064***	0.056***	0.063***	0.062***	1.06	0.01
		(0.016)	(0.015)	(0.015)	(0.015)		
	FIXED EFFECTS						
	violation vintage	Included	Included	Included	Included		

Table 4.3: Regression Results (continued)

violation functional sub-category	Included	Included	Included	Included
audit year	Included	Included	Included	Included
country			Included	Included
industry				Included
RANDOM EFFECTS				
audit	Included	Included	Included	Included
factory	Included	Included	Included	Included

Notes: *p<0.10, **p<0.05, ***p<0.01; N = 72,983 observations from 8,616 audits of 2,626 factories. Model is estimated using a multilevel mixed-effects logistic regression model with standard errors clustered at the factory. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model. The odds ratios (OR) presented in column 5 are calculated by exponentiating the coefficients in column 5. The average marginal effects (AME) presented in column 6 reflect the marginal predicted mean of violation resolution probability from the results in column 4.

The results of several control variables in the full mode found in Table 3.3, column 4, provide some insight into different factors that are correlated with violation resolution odds. First, we do not find that prior unannounced audits (*unannounced (prior audit)*) are significantly correlated with lower odds of violation resolution, which is inconsistent with prior literature that finds this significant positive correlation (Short, Toffel, and Hugill 2020). This difference could be driven by our different measures of improvement and could be explored further in future work. I also find that the odds of violation resolution are significantly lower (21% lower) when focal audits are unannounced (variable *unannounced*) versus when audits are pre-announced, possibly because factories receiving unannounced audits have less lead time to gather sufficient evidence which demonstrates violation closure. The results also demonstrate that the odds of violation closure are slightly lower (1% lower) for factories with more violations in the prior audit (variable *number of recorded violations (prior audit)*). This could happen because factories with more prior violations must spread their attention (and financial resources) over more violations that require closing, which leaves them with fewer attentional resources to spend on each individual violation's resolution. The odds of violation resolution are 16% lower for *major* violations than for minor ones, which is consistent with the notion that major violations are harder to resolve. Finally, odds of violation resolution increase the more factories are audited (variable *audit sequence*), on average 6%, which suggests factories accumulate knowledge on how to resolve violations the more they are audited.

Table 3.3, column 4 also presents results for the three hypothesized variables. The positive and statistically significant coefficient on *percent supplied (prior audit)* indicates that the odds of violation resolution increases when a larger portion of the supplier's output is supplied to the brand, which supports hypothesis 1. Specifically, these results suggest that a ten percentage point increase in *percent supplied (prior audit)* is associated with a small 2.2% increase in the odds of violation resolution in the focal audit, or a one standard deviation increase in percent supplied (prior audit) is associated with a 6.1% increase in violation resolution odds, all else equal. In terms of average marginal effects, these results suggest that increasing *percent supplied (prior audit)* from 0% to 100% is associated with a 7.5% increase in the average predicted probability of violation resolution in the focal audit from 44.0% to 47.3%. The odds ratio corresponding to the coefficient on *number of alternative suppliers (prior audit)*, a variable modeled using a natural log, is 1.02 and is significant at the 1% level, which supports hypothesis 2. This result indicates that increasing the natural log of number of alternative suppliers in the brand's procurement portfolio that are in the supplier's industry by one corresponds to a small 2% increase in violation resolution odds. Put another way, a 1 standard deviation increase in this variable is associated with a 5.4% increase in resolution odds. In terms of average marginal effects, these results suggest that roughly doubling the *number of alternative suppliers (prior* *audit)* from unlogged values of 678 to 1,260 is associated with a small increase in the average predicted probability of violation resolution in the focal audit from 44.9% to 45.1%.[91] These results support hypothesis 1 and 2, although the effect sizes are quite small in magnitude.

Finally, the negative and statistically significant coefficient on *second-party (prior audit)* opposes the third hypothesis, that factories would be more likely to resolve violations following a second-party audit. Instead, this coefficient suggests that factories are associated with a 9% decrease in violation resolution odds following a second-party audit, all else equal. While surprising, this result could be driven by the following scenario: second-party auditors tend to record more violations in audits (Ibanez et al. 2023), so second-party auditors may spend more of their limited auditing time toward identifying new issues at the factory and spend less time communicating how to resolve active issues to factory management.

Robustness tests. To ensure that main model results are not driven by the less frequent occurrence of minor violations (15% of observations), I re-estimated the main model using only major violations. These results can be found in Table E-1, and their results are quite similar to the main results displayed in Table 3.3, column 4. This provides evidence that my main results are robust to the omission of minor violations and also suggests that prior audit switching costs influence focal audit resolution odds of more severe violations.

Specifically, in column 1 of Table E-1 I find that a 10% increase in *percent supplied (prior audit)* increases the odds of violation resolution by 2.2%. The odds ratio corresponding to the coefficient on *number of alternative suppliers (prior audit)* is 1.03 and is also significant at

---

[91] This was found by finding the average predicted probability of violation resolution when logged values of *number of alternative suppliers (prior audit)* approximately equal 6.52 (the 25th percentile) and 7.14 (the 50th percentile). These logged values correspond to unlogged values of 678 and 1,260 by exponentiating the logged values and subtracting 1.

the 1% level, indicates that increasing the natural log of the number of alternative suppliers in the brand's portfolio by 1 corresponds to a small increase in violation resolution odds (by 3%). Interestingly, the negative coefficient on *second-party (prior audit)* becomes non-significant with the omission of minor violations, which suggests that the reduction in violation resolution odds when the prior audit was performed by second-party auditors may be driven by supplier's future compliance decisions around minor violations.

The main model specification I utilize in this analysis performs a cross-factory analysis that does not explicitly hold constant factory-specific violation resolution odds, which could be influenced by unobserved variables (such as the perceptions and attitudes of factory management toward workplace hygiene and worker safety). I therefore re-estimated a model using a conditional logistic regression (conditional on the factory) that includes factory fixed effects and cluster standard errors at the factory. While doing so introduces right censoring of the data (e.g., a factory is dropped from this sample if no violation is resolved before the data panel ends in 2017), this model performs a within-factory analysis and thus holds constant the factory-specific average propensity to resolve violations.[92]

Table E-2 presents these results. I find similar levels of statistical significance for hypotheses 1 and 2. Specifically, I find that a 10% increase in *percent supplied (prior audit)* increases the odds of violation resolution by 3.4%, which is about 1 percentage point higher than the main model results for this variable. The odds ratio corresponding to the coefficient on *number of alternative suppliers (prior audit)* is 1.05 and is also significant at the 1% level, indicates that increasing the natural log of number of alternative supplier's in the brand's

[92] To achieve model convergence with the inclusion of factory fixed effects, it was necessary to omit the 92 functional sub-category violation categories and instead include 24 fixed effects for each combination of violation severity (major or minor) and violation parent functional category (of which there are 12 categories).

portfolio in the supplier's industry by one corresponds to a small increase in violation resolution odds by 5%, which is more than double the increase in violation resolution odds for this variable found in the main results. Together these results suggest that potential factory-specific heterogeneity in violation resolution odds from unobservable time-stable characteristics do not materially impact the main model results.

One assumption I make in the construction of *number of alternative suppliers (prior audit)* is that using product codes to identify a supplier's industry results in a sample whose industries may not be a reasonable representation of the global supplier population. An ideal test of this assumption would be to compare the share of suppliers by country, industry, and year in my data to that in an external data source. However, because I have access to product code numbers, but not the names of these coded products, I cannot perform this comparison. Instead, I separately constructed an industry measure for all suppliers in my data using three sources of data. I then found that model results that include an alternative measure of *number of alternative suppliers (prior audit)* constructed using the alternative supplier industry category yields similar results to the main model H2 results (see Appendix Table I-1), which indicates a positive correlation between *number of alternative suppliers (prior audit)* using either industry categorization. I then compare, in Table I-2, the share of 2012 suppliers in China, the country with the greatest number of suppliers and audits, in TrendyStyle's database to the share of 2012 enterprises in China by industry as reported by China Data Online. [93] The similarities in supplier share by industry across both data sets provides some evidence that both industry categories are a

[93] Enterprise reflects a number of different company ownership categories: those that are state-owned, privately owned, cooerpatives, collectively owned, those that are shareholding enterprises, and those with funds from foreign entities.

reasonable representation of the global supplier population. This additional analysis is described in greater detail in Appendix I.

In Table E-3, I also re-estimate the main model using an alternative measure to test H2. I relax the time constraint on the construction of *number of alternative suppliers (prior audit)* by constructing a measure which equals the number of other suppliers in the focal supplier's industry that the brand audits within the prior two years that precedes the prior audit, and taking the natural log of this value to reduce skew. Allowing suppliers to be counted in this measure if they were audited by the brand in the prior two years versus the prior year (which is the time constraint of the measure included in the main model) does not materially affect results.

The mechanisms theorized above should not only lead to higher odds of violation resolution in the focal audit, but should also lead to greater overall compliance in the focal audit in terms of fewer audit violations. To explore whether this is true, I estimate a Poisson model at the factory-audit level that predicts the number of focal audit violations, and that includes all hypothesized variables and all audit-level control variables included in the main model.[94] Column 1 of Table E-4 presents the results from this Poisson model and includes factory fixed effects (performing a within-factory analysis).[95] Column 2 of Table E-4 employs a specification with random effects at the factory (performing an across-factory analysis).[96] The negative and statistically significant coefficients on *percent supplied (prior audit)* and *number of alternative suppliers (prior audit)* using either a fixed effect (column 1) or random effect (column 2)

---

[94] Because I estimate an audit-level regression, I omit all violation-level variables and fixed effects, such as violation severity and violation functional sub-category.
[95] This model does not include the country or industry fixed effects that are in the main model due to the inclusion of factory fixed effects.
[96] The inclusion of factory random effects in Column 2 assumes that variation across factories is random and is uncorrelated with explanatory variables.

specification reinforce the mechanisms theorized for H1 and H2. That is, these results suggest that having higher supplier switching costs (lower buyer switching costs) are associated with greater focal audit compliance in terms of fewer focal audit violations.[97]

CONCLUSION

This study reveals important insights about how buyers can use their economic relationship with suppliers as consumers to affect improvement in supplier working conditions. Specifically, I identify two aspects of the buyer's economic relationship with suppliers that I theorize raise (lower) supplier (buyer) switching costs and that better incentivize suppliers to resolve violations. First, factories with higher costs of switching to a different buyer, as measured by the percent output they supply to this buyer at the time of the prior audit, are more likely to resolve a given violation in the next audit. Second, factories that operate in industries where the buyer has lower costs of switching to a different factory supplier, as measured by the number of alternative suppliers the in the buyer's auditing portfolio within a year of the prior audit, are also more likely to resolve a given violation in the next audit. Against my hypothesis, I find that factories are less likely to resolve violations after a second-party audit.

This study primarily contributes to the literature on supply chain monitoring quality in a few ways. A central question in this literature is whether monitoring programs are effective, or whether they reflect "window dressings" that are de-coupled from actual supplier practices (Boiral 2007; Bromley and Powell 2012). The research in this literature stream that explores the

[97] The main model results for H3 are counter to what I hypothesize, that factories are more likely to improve following second-party audits. Specifically, the negative and statistically significant coefficient on *second-party (prior audit)* in Table 3.3, column 4, indicates that the odds of violation resolution are lower following a second-party audit (versus third-party audit), all else equal. The positive and statistically significant coefficient on *second-party (prior audit)* in Table E-4 is consistent with the results for this variable in Table 3.3. That is, factories are less compliant in the focal audit, in terms of having more violations, following a second-party audit.

effectiveness of such monitoring programs tend to focus on identifying conditions under which auditors record more violations. While this focus is crucial to help guide managers in their designing effective monitoring programs, so too is understanding whether factories actually remedy those violations, and under what conditions they are more likely to resolve them. A few scholars have identified important institutional conditions (Seidman 2007) and supplier management practices (Bird, Short, and Toffel 2019) that lead to greater improvement in factory working conditions, and this study adds to that literature by exploring how economic relationships between buyers and suppliers affect supplier improvement.

In identifying important characteristics of economic relationships between buyers and suppliers, this study provides two potential levers that companies can pull to incentive greater supplier improvement in their auditing programs. First, these results suggest that companies should consider rewarding more compliant suppliers by offering larger future purchase orders, as suggested by non-profit organizations (Fair Labor Association 2015). Second, these results also suggest that companies should diversify their purchase orders from many suppliers in each industry they procure from, as this practice tends to lead to greater improvement in working conditions. Taking up both suggestions simultaneously may not be feasible, as it would result in shifting procurement toward suppliers that become more compliant *and* shifting procurement to more suppliers in the focal supplier's industry. Therefore, companies could consider using these strategies sequentially, by first offering a reward of larger future purchase orders for suppliers with improved compliance, and then by increasing the number of suppliers they procure from in each industry if factories do not respond to the first incentive.

There are several limitations of this study. First, both variables hypothesized in hypotheses 1 and 2 are fundamentally endogenous in the sense that TrendyStyle consciously decides the level of procurement volume from any given supplier (H1) and the number of suppliers to procure from in any given industry (H2). While I include several additional tests to ensure that TrendyStyle's procurement decisions are plausibly unrelated to audit outcomes, it is possible that my results suffer from omitted variable bias if an unobserved variable affects both procurement decisions and audit outcomes. Therefore, the results presented here should be interpreted as detected statistical correlations between hypothesized variables and outcome measure, and not as strong causal links between variables.

Second, it is possible that the mechanisms theorized above affect not only the compliance decisions that suppliers make between the prior and focal audit, but also affect the compliance decisions that suppliers make at the time of the prior audit. This could bias my results if the effect of the mechanisms on compliance in the prior audit leads to unobserved differences in the set of prior audit violations that are present at different factories which are more or less impacted by the proposed mechanism in ways that affects the hypothesized variables and violation resolution propensity. For example, if factories that, in the prior audit, face greater cost of switching buyers, that greater switching cost may mean that the violations that are present in the prior audit are on average easier to resolve than violations at factories facing lower costs of switching buyers. While my comparison of violation characteristics across suppliers theorized to have lower or higher switching costs did not reveal any significant differences in observed violation characteristics, it is possible that there may be violation differences I cannot observe that drive prior audit compliance and focal violation resolution odds. Future research investigating violation closure could survey factory suppliers to understand what violation characteristics tend to hinder their ability to resolve them and, if necessary, create a measure using survey responses to better account for these unobserved factors that may influence violation closure.

Third, the data I employ come from audits of one buyer, which naturally presents a few limitations. It is possible that this study's results are buyer specific. Future research should replicate this analysis using data from multiple buyers. Additionally, data limitations may have limited my construction of certain variables, which could lead to measurement error. When constructing *percent supplied (prior audit),* I do not know how many other buyers the supplier factories supply to, which could impact supplier switching costs. Similarly, when constructing *alternative number of suppliers (prior audit)*, I can only observe other suppliers in the focal supplier's industry that the buyer is already procuring from and is auditing. A more refined version of this measure would consider both suppliers in the same industry as the focal supplier that the buyer is already auditing and other suppliers that operate in the same industry but that are not in the buyer's procurement portfolio.

There are also future directions of research that warrant extended discussion. One dimension of the audit report that suppliers may use to resolve violations is the corrective action they are provided with after the recording of a violation. This feedback for each violation should describe what specific actions the supplier should take to resolve the issue. In Appendix H, I describe specific ways that future work could use auditor comments, specifically how informative the comment is or the comments sentiment, to predict violation resolution.

While firms that utilize global supply chains are increasingly deploying monitoring programs to monitor the working conditions of their suppliers, it is not understood whether and how these programs are effective in their incentivizing suppliers to resolve active workplace violations. My study of supply chain audit outcomes at thousands of globally-located factories suggests one important factor – the firm's economic relationship with suppliers – can incentive improvement in supplier working conditions. This study also contributes to the literature on supply chain monitoring quality, specifically to the subset of this literature on supplier compliance improvement, and highlights potential strategies to improve supplier working conditions.

REFERENCES

Abolfathi, N., Santamaria, S., & Williams, C. (2022). How does firm scope depend on customer switching costs? Evidence from mobile telecommunications markets. *Management Science*, *68*(1), 316-332.

AFL-CIO. (2013). *Responsibility Outsourced: Social Audits, Workplace Certification and Twenty Years of Failure to Protect*. http://www.aflcio.org/Learn-About-Unions/Global-Labor-Movement/Responsibility-Outsourced-Report (accessed December 18, 2021).

Al-Thuneibat, A.A., Al Issa, R.T.I., & Baker, R.A.A. (2011). Do audit tenure and firm size contribute to audit quality? *Managerial Auditing Journal, 26*(4), 317–334.

Ammenberg, J., Wik, G., & Hjelm, O. (2001). Auditing external environmental auditors—Investigating how ISO 14001 is interpreted and applied in reality. *Eco-Management and Auditing: The Journal of Corporate Environmental Management, 8*(4), 183–192.

Anderson, E., & Schmittlein, D.C. (1984). Integration of the sales force: An empirical examination. *RAND Journal of Economics, 15*(3), 385–395.

Ball, G., Siemsen, E., & Shah, R. (2017). Do plant inspections predict future quality? The role of investigator experience. *Manufacturing Service Operations Management*, *19*(4), 534-550.

Bandura, A. (1991). Social theory of self-regulation. *Organizational Behavior and Human Decision Processes*, *50*(2), 248–287.

Bandura, A., & Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of Personality and Social Psychology*, *45*(5), 1017–1028.

Barkemeyer, R. (2009). Beyond Compliance—Below Expectations? CSR in the Context of International Development. *Business Ethics: A European Review, 18*(3), 273–289.

Baron, R. A. (1988). Negative effects of destructive criticism: Impact on conflict, self-efficacy, and task performance. *Journal of Applied Psychology*, *73*(2), 199–207.

Baron, R. A. (1990). Counteracting the effects of destructive criticism: The relative efficacy of four interventions. *Journal of Applied Psychology*, *75*(3), 235–245.

Baron, R. A. (1993). Criticism (informal negative feedback) as a source of perceived unfairness in organizations: Effects, mechanisms, and countermeasures. In R. Cropanzano (Ed.), Justice in the workplace: Approaching fairness in human resource management (pp. 155–170). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Barrero, J.M., Bloom, N., & Davis, S.J. (2021). Why Working from Home Will Stick. NBER Working Paper No. 28731, April 2021.

Bartley, T., & Childs, C. (2014). Shaming the corporation: The social production of targets and the anti-sweatshop movement. *American Sociological Review, 79*(4), 653–679.

Bavafa, H., Hitt, L.M., & Terwiesch, C. (2018). The Impact of E-Visits on Visit Frequencies and Patient Health: Evidence from Primary Care. *Management Science. 64*(12), 5461–5480.

Bell, D.R., Gallino, S., & Moreno, A. (2018). Offline Showrooms in Omnichannel Retail: Demand and Operational Benefits. *Management Sci*ence, *64*(4), 1629–1651.

Bennett, V.M., Pierce, L., Snyder, J.A., & Toffel, M.W. (2013). Customer-driven misconduct: How competition corrupts business practices. *Management Science, 59*(8), 1725–1742.

Bettinger, E.P., Fox, L., Loeb, S., & Taylor, E.S. (2017). Virtual Classrooms: How Online College Courses Affect Student Success. *American Economic Review, 107*(9), 2855–2875.

Bird, Y., Short, J.L. & Toffel, M.W. (2019). Coupling labor codes of conduct and supplier labor practices: The role of internal structural conditions. *Organization Science, 30*(4), 847–867.

Bloom, N., Liang, J., Roberts, J., & Ying, Z.J. (2015). Does Working from Home Work? Evidence from a Chinese Experiment. *Quarterly Journal of Economics, 130*(1), 165–218.

Bohannon, L.S., Herber, A.M., Pelz, J.B., & Rantanen, E.M. (2013). Eye contact and video-mediated communication: A review. *Displays*, *34*(2), 177-185.

Boiral, O. (2003). ISO 9000: Outside the iron cage. *Organization Science, 14*(6), 720–737.

Boiral, O. (2007). Corporate Greening through ISO 14001: A Rational Myth? *Organization Science 18*(1), 127–146.

Boiral, O., & Gendron, Y. (2011). Sustainable development and certification practices: Lessons learned and prospects. *Business Strategy and the Environment, 20*(5), 331–347.

Boone, J.P., Khurana, I.K., & Raman, K.K. (2012). Audit market concentration and auditor tolerance for earnings management. *Contemporary Accounting Research, 29*(4), 1171–1203.

Borgatti, S.P., & Cross, R. (2003). A Relational View of Information Seeking and Learning in Social Networks. *Management Science, 49*(4), 432–445.

Bradach, J.L. (1997). Using the plural form in the management of restaurant chains. *Administrative Science Quarterly, 42*(2), 276–303.

Bradach, J.L., & Eccles, R.G. (1989). Price, authority, and trust: From ideal types to plural forms. *Annual Review of Sociology, 15*(1), 97–118.

Braithwaite, J., & Makkai, T. (1991). Testing an expected utility model of corporate deterrence. *Law & Society Review, 25*(1), 7-40.

Brass, D.J., Butterfield, K.D., & Skaggs, B.C. (1998). Relationships and unethical behavior: A social network perspective. *Academy of Management Review, 23*(1), 14–31.

Bromley, P., & Powell, W.W. (2012). From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World. *Academy of Management Annals, 6*(1), 483–530.

Brush, T. H., Dangol, R., & O'Brien, J. P. (2012). Customer capabilities, switching costs, and bank performance. *Strategic Management Journal*, *33*(13), 1499-1515.

Buell, R.W., Campbell, D., & Frei, F. (2010). Are Self-Service Customers Satisfied or Stuck? *Production Operations Management, 19*(6), 679–697.

Cameron, A.C., & Trivedi, P.K. (1998). *Regression analysis of count data*, New York: Cambridge University Press.

Campbell, D., & Frei, F. (2010). Cost Structure, Customer Profitability, and Retention Implications of Self-Service Distribution Channels: Evidence from Customer Behavior in an Online Banking Channel. *Management Science, 56*(1), 4–24.

Carcello, J.V., & Nagy, A.L. (2004). Audit firm tenure and fraudulent financial reporting. *Auditing: A Journal of Practice & Theory, 23*(2), 55–69.

Carnahan, S., & Somaya, D. (2013). Alumni effects and relational advantage: The impact on outsourcing when a buyer hires employees from a supplier's competitors. *Academy of Management Journal*, *56*(6), 1578-1600.

Causholli, M., Chambers, D.J., & Payne, J.L. (2014). Future nonaudit service fees and audit quality. *Contemporary Accounting Research, 31*(3), 681–712.

Chebat, J. C., Davidow, M., & Borges, A. (2011). More on the role of switching costs in service markets: A research note. *Journal of Business Research*, *64*(8), 823-829.

Chen, P.Y., & Hitt, LM. (2002). Measuring switching costs and their determinants in Internet enabled business: a study of the online brokerage industry. *Information Systems Research*, *13*(3), 255-274.

Choudhury, P.R., Foroughi, C., & Larson, B. (2021). Work-from-anywhere: The productivity effects of geographic flexibility. *Strategic Management Journal, 42*(4), 655– 683.

Chow, G.C. (1960). Tests of Equality between Sets of Coefficients in Two Linear Regressions. *Econometrica*, *28*, 591-605

Chugh, D., & Bazerman, M.H. (2007). Bounded awareness: What you fail to see can hurt you. *Mind and Society, 6*(1), 1–18.

Coffee, J. C., Jr. (2004). Gatekeeper failure and reform: The challenge of fashioning relevant reforms. *Boston University Law Review, 84*(2), 301–364.

Cramton, C.D. (2001). The Mutual Knowledge Problem and Its Consequences for Dispersed Collaboration. *Organization Science, 12*(3), 346–371.

Cressey, D.R., & Moore, C.A. (1983). Managerial values and corporate codes of ethics. *California Management Review, 25*(4), 53–77.

Cui, R., Gallino, S., Moreno, A., & Zhang, D. J. (2018). The operational value of social media information. *Production and Operations Management*, *27*(10), 1749-1769.

Dao, M., Raghunandan, K., & Rama, D.V. (2012). Shareholder voting on auditor selection, audit fees, and audit quality. *Accounting Review, 87*(1), 149–171.

Darnall, N., Ji, H., & Vázquez-Brust, D.A. (2018). Third-party certification, sponsorship, and consumers' ecolabel use. *Journal of Business Ethics, 150*(4), 953–969.

DeAngelo, L.E. (1981). Auditor size and audit quality. *Journal of Accounting and Economics, 3*(3), 183–199.

Distelhorst, G., Locke, R.M., Pal, T., & Samel, H. (2015). Production goes global, compliance stays local: Private regulation in the global electronics industry. *Regulation & Governance, 9*(3), 224–242.

Distelhorst, G., Hainmueller, J., & Locke, R.M. (2017). Does Lean Performance Improve Labor Standards? Management and Social Performance in the Nike Supply Chain. *Management Science, 63*(3), 707-728.

Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. *Quarterly Journal of Economics, 128*(4), 1499–1545.

Dutta, S., Bergen, M., Heide, J.B, & John, G. (1995). Understanding dual distribution: The case of reps and house accounts. *Journal of Law, Economics, & Organization, 11*(1), 189–204.

Fan, M., Ke, B., Wu, W., Xia, L., & Xin, Q. (2020). Should home country auditors be allowed to audit cross-listed firms: Evidence from China. SSRN Working Paper, https://ssrn.com/abstract=3588185.

Fair Labor Association. (2015). 2015 Annual Report, Fair Labor Association.

Fayol, H. (1967). General and Industrial Management. London: Pitman.

Fedor, D., Davis, W., Maslyn, J., & Mathieson, K. (2001). Performance improvement efforts in response to negative feedback: The roles of source power and recipient self-esteem. *Journal of Management*, *27*(1), 79–97.

Francis, J.R., & Wilson, E.R. (1988). Auditor changes: A joint test of theories relating to agency costs and auditor differentiation. *Accounting Review, 64*(4), 663–682.

Gómez, J., & Maícas, J. P. (2011). Do switching costs mediate the relationship between entry timing and performance?. *Strategic Management Journal*, *32*(12), 1251-1269.

Gibson, C.B., & Gibbs, J.L. (2006). Unpacking the Concept of Virtuality: The Effects of Geographic Dispersion, Electronic Dependence, Dynamic Structure, and National Diversity on Team Innovation. *Administrative Science Quarterly, 51*(3), 451–495.

Gray, W.B., & Shadbegian, R. (2005). When and why do plants comply? Paper mills in the 1980s. *Law & Policy, 27*(2), 238-261.

Greene, W.H. (2008). Econometric Analysis. 6th Edition, Pearson Prentice Hall, Upper Saddle River.

Greenberg, J., & Mollick, E. (2017). Activist choice homophily and the crowdfunding of female founders. *Administrative Science Quarterly, 62*(2), 341–374.

Griffin, J.M., & Tang, D.Y. (2011). Did credit rating agencies make unbiased assumptions on CDOs? *American Economic Review, 101*(3), 125–130.

Grzybowski, L. (2008). Estimating switching costs in mobile telephony in the UK. *Journal of Industry, Competition and Trade*, 8(2), 113-132.

Gul, F.A., Wu, D., & Yang, Z. (2013). Do individual auditors affect audit quality? Evidence from archival data. *Accounting Review, 88*(6), 1993–2023.

He, D., & Nickerson, J.A. (2006). Why do firms make and buy? Efficiency, appropriability, and competition in the trucking industry. *Strategic Organization, 4*(1), 43–69.

Heide, J. B., & Weiss, A. M. (1995). Vendor consideration and switching behavior for buyers in high-technology markets. *Journal of Marketing*, *59*(3), 30-43.

Heide, J.B. (2003). Plural governance in industrial purchasing. *Journal of Marketing, 67*(4), 18–29.

Heide, J.B., Kumar, A., & Wathne, K.H. (2014). Concurrent sourcing, governance mechanisms, and performance outcomes in industrial value chains. *Strategic Management Journal, 35*(8), 1164–1185.

Hinds, P.J., & Mortensen M. (2005). Understanding Conflict in Geographically Distributed Teams: The Moderating Effects of Shared Identity, Shared Context, and Spontaneous Communication. *Organization Science, 16*(3), 290–307.

Huang, T., Chang, H., & Chiou, J. (2016). Audit market concentration, audit fees, and audit quality: Evidence from China. *Auditing: A Journal of Practice & Theory, 35*(2), 121–145.

Iacus, S.M., King, G. and Porro, G. (2008). Matching for causal inference without balance checking. Available at SSRN 1152391.

IAF, ILAC, & ISO. (2021). Use of Remote Techniques supported by IAF/ILAC/ISO Survey. Retrieved March 9, 2023, https://ilac.org/latest_ilac_news/use-of-remote-techniques-supported-by-iaf-ilac-iso-survey/

Ibanez, M.R., Palmarozzo, A., Short, J.L., & Toffel, M.W. (2022). Second- vs. Third-Party Audit Quality: Evidence from Global Supply Chain Monitoring. Working Paper, Harvard Business School, Boston.

Ibanez, M.R., & Toffel, M.W. (2020). How scheduling can bias quality assessment: Evidence from food-safety inspections. *Management Science, 66*(6), 2396–2416.

ISO 19011. (2018). Guidelines for Auditing Management Systems (ISO 19011:2018). International Organization for Standardization (ISO): Geneva, Switzerland.

Jensen, M.C., & Meckling, W.H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360.

Jia, N. (2018). The "make and/or buy" decisions of corporate political lobbying: Integrating the economic efficiency and legitimacy perspectives. *Academy of Management Review, 43*(2), 307–326.

Jiang, J.X., Stanford, M.H., & Xie, Y. (2012). Does it matter who pays for bond ratings? Historical evidence. *Journal of Financial Economics, 105*(3), 607–621.

Jin, G. Z., & Lee, J. (2018). A tale of repetition: Lessons from Florida restaurant inspections. *Journal of Law and Economics, 61*(1), 159–188.

John, G., & Weitz, B.A. (1988). Forward integration into distribution: An empirical test of transaction cost analysis. *Journal of Law, Economics, and Organization, 4*(2), 337–355.

Kang, Y., Cai, Z., Tan, C. W., Huang, Q., & Liu, H. (2020). Natural language processing (NLP) in management research: A literature review. *Journal of Management Analytics*, *7*(2), 139-172.

Khalil, F., & Lawarrée, J. (2006). Incentives for corruptible auditors in the absence of commitment. *Journal of Industrial Economics, 54*(2), 269–291.

King, B., & McDonnell, M. (2015). Good firms, good targets: The relationship among corporate social responsibility, reputation, and activist targeting. In K. Tsutsui & A. Lim (Eds.), *Corporate social responsibility in a globalizing world* (pp. 430–454). New York: Cambridge University Press.

Kinney, W.R. Jr., Palmrose, Z., & Scholz, S. (2004). Auditor independence, non-audit services, and restatements: was the U.S. government right? *Journal of Accounting Research, 42*(3), 561–588.

Klein, B., Crawford, R.G., & Alchian, A.A. (1978). Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics, 21*(2), 297–326.

Klemperer, P. (1995). Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *The Review of Economic Studies*, *62*(4), 515-539.

Klotz, D., & Chatterjee, K. (1995). Dual sourcing in repeated procurement competitions. *Management Science, 41*(8), 1317–1327.

Kofman, F., & Lawarrée, J. (1993). Collusion in hierarchical agency. *Econometrica, 61*(3), 629–656.

Kouakou, D., Boiral, O., & Gendron, Y. (2013). ISO auditing and the construction of trust in auditor independence. *Accounting, Auditing & Accountability Journal, 26*(8), 1279–1305.

Kraakman, R.H. (1986). Gatekeepers: The anatomy of a third-party enforcement strategy. *Journal of Law, Economics, & Organization, 2*(1), 53–104.

Krogh, G.V., Roos, J., & Slocum, K. (1994). An essay on corporate epistemology. *Strategic Management Journal, 15*(S2), 53–71.

Kumar, J.A., & Chakrabarti, A. (2012). Bounded awareness and tacit knowledge: Revisiting Challenger disaster. *Journal of Knowledge Management, 16*(6), 934–949.

Lebaron, G., & Lister, J. (2015). Benchmarking global supply chains: The power of the "ethical audit" regime. *Review of International Studies, 41*(05), 905–924.

Lennox, C.S. (2014). Auditor tenure and rotation. In D. Hay, W. R. Knechel, & M. Willekens (Eds.), *The Routledge companion to auditing*, *1st edition* (pp. 89-106). New York: Routledge.

Lennox, C.S., Wu, X., & Zhang, T. (2014). Does mandatory rotation of audit partners improve audit quality? *The Accounting Review, 89*(5), 1775–1803.

Levitt, A. (2000). Renewing the covenant with investors. Speech delivered to Center for Law and Business, New York University, May 10, 2000, US Securities and Exchange Commission. https://www.sec.gov/news/speech/spch370.htm, accessed August 2020.

Liu, A. H. (2006). Customer value and switching costs in business services: developing exit barriers through strategic value management. *Journal of Business & Industrial Marketing*, *21*(1), 30-37.

Ljungqvist, A., Marston, F., Starks, L.T., Wei, K.D., & Yan, H. (2007). Conflicts of interest in sell-side research and the moderating role of institutional investors. *Journal of Financial Economics, 85*(2), 420–456.

Locke, R.M., Amengual, M., & Mangla, A. (2009). Virtue out of necessity? Compliance, commitment, and the improvement of labor conditions in global supply chains. *Politics & Society, 37*(3), 319–351.

Locke, R.M., Qin, F., & Brause, A. (2007). Does monitoring improve labor standards? Lessons from Nike. *ILR Review, 61*(1), 3–31.

Lundberg, S., & Stearns, J. (2019). Women in economics: Stalled progress. *Journal of Economic Perspectives, 33*(1), 3–22.

Lytton, T.D., & McAllister, L.K. (2014). Oversight in private food safety auditing: Addressing auditor conflict of interest. *Wisconsin Law Review, 2014*(2), 289–336.

Macher, J.T., Mayo, J.W., & Nickerson, J.A. (2011). Regulator Heterogeneity and Endogenous Efforts to Close the Information Asymmetry Gap. *Journal of Law & Econom*ics, *54*(1), 25–54.

Macher, J.T., & Richman, B.D. (2008). Transaction cost economics: An assessment of empirical research in the social sciences. *Business and Politics, 10*(1), 1–63.

Malesky, E. J., & Mosley, L. (2018). Chains of love? Global production and the firm-level diffusion of labor standards. *American Journal of Political Science*, *62*(3), 712-728.

Manik, J.A., & Yardley, J. (2012, December 17). Bangladesh finds gross negligence in factory fire. *New York Times*. https://www.nytimes.com/2012/12/18/world/asia/bangladesh-factory-fire-caused-by-gross-negligence.html, accessed May 2021.

Markell, D.L., & Glicksman, R.L. (2014). A holistic look at agency enforcement. *North Carolina Law Review, 93*(1), 1–18.

Mayer, K.J. (2006). Spillovers and governance: An analysis of knowledge and reputational spillovers in information technology. *Academy of Management Journal, 49*(1), 69–84.

Mayhew, B.W., & Pike, J.E. (2004). Does investor selection of auditors enhance auditor independence? *The Accounting Review, 79*(3), 797–822.

McAllister, L.K. (2012). Regulation by third-party verification. *Boston College Law Review, 53*(1), 1–32.

McBarnet, D. (2007). Corporate social responsibility beyond law, through law, for law: The new corporate accountability. In D. McBarnet, A. Voiculescu, & T. Campbell (Eds.), *The new corporate accountability: Corporate social responsibility and the law* (pp. 9–56). Cambridge, UK: Cambridge University Press.

Meidinger, E. (2019). Governance interactions in sustainable supply chain management. In S. Wood, R. Schmidt, E. Meidinger, B. Eberlein, & K.W. Abbott (Eds.), *Transnational Business Governance Interactions: Advancing Marginalized Actors and Enhancing Regulatory Quality* (pp. 52-76). Cheltenham, UK, and Northampton, MA: Edward Elgar.

Merchant, B (2017, June 18). Life and death in Apple's forbidden city. *The Guardian (UK)*. https://www.theguardian.com/technology/2017/jun/18/foxconn-life-death-forbidden-city-longhua-suicide-apple-iphone-brian-merchant-one-device-extract, accessed May 2021.

Metzger, M., Dalton, D.R., & Hill, J.W. (1993). The organization of ethics and the ethics of organizations: The case for expanded organizational ethics audits. *Business Ethics Quarterly, 3*(1), 27–43.

Meyer, J.W., & Rowan, R. (1977). Institutionalized Organizations: Formal Structure as Myth and Ceremony. *American Journal of Sociology,* 83(2), 340–363.

Montiel, I., Husted, B.W., & Christmann, P. (2012). Using private management standard certification to reduce information asymmetries in corrupt environments. *Strategic Management Journal, 33*(9), 1103–1113.

Moore, D.A., Tetlock, P.E., Tanlu, L., & Bazerman, M.H. (2006). Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *Academy of Management Review, 31*(1), 10–29.

Newton, N.J., Wang, D., Wilkins, M.S. (2013). Does a lack of choice lead to lower quality? Evidence from auditor competition and client restatements. *Auditing: A Journal of Practice & Theory, 32*(3), 31–67.

Nickerson, J.A., & Silverman, B.S. (2003). Why aren't all truck drivers owner-operators? Asset ownership and the employment relation in interstate for-hire trucking. *Journal of Economics & Management Strategy, 12*(1), 91–118.

Norton, S.W. (1988). An empirical look at franchising as an organizational form. *Journal of Business, 61*(2), 197–218.

O'Rourke, D. (2003). Outsourcing regulation: Analyzing nongovernmental systems of labor standards and monitoring. *Policy Studies Journal, 31*(1), 1–29.

Oka, C. (2010). Accounting for the Gaps in Labour Standard Compliance: The Role of Reputation-conscious Buyers in the Cambodian Garment Industry. *European Journal of Development Research, 22*(1), 59–78.

Oka, C. (2015). Improving Working Conditions in Garment Supply Chains: The Role of Unions in Cambodia. *British Journal of Industrial Relations, 54*(3), 647-672.

Parmigiani, A. (2007). Why do firms both make and buy? An investigation of concurrent sourcing. *Strategic Management Journal, 28*(3), 285–311.

Peñarroja, V., Orengo, V., Zornoza, A., & Hernández, A. (2013). The effects of virtuality level on task-related collaborative behaviors: The mediating role of team trust. *Computers and Human Behavior*, *29*(3), 967-974.

Pierce L., & Toffel, M.W. (2013). The role of organizational scope and governance in strengthening private monitoring. *Organization Science, 24*(5), 1558–1584.

Porter, M. E. (2008). The five competitive forces that shape strategy. *Harvard Business Review*, *86*(1), 78.

Power, M. (1997). *The audit society: Rituals of verification*. Oxford, UK: Oxford University Press.

Power, D. and Terziovski, M. 2007. Quality audit roles and skills: Perceptions of non-financial auditors and their clients. *Journal of Operations Management,* 25(1), 126-147.

Prakash, A., & Potoski, M. (2007). Collective action through voluntary environmental programs: A club theory perspective. *Policy Studies Journal, 35*(4), 773–792.

Puranam, P., Gulati, R., & Bhattacharya, S. (2013). How much to make and how much to buy? An analysis of optimal plural sourcing strategies. *Strategic Management Journal, 34*(10), 1145–1161.

Rajan, B., Tezcan, T., Seidmann, A. (2019). Service Systems with Heterogeneous Customers: Investigating the Effect of Telemedicine on Chronic Care. *Management Science*, *65*(3), 1236–1267.

Reid, D.J., & Reid, F.J. (2005). Online focus groups: An in-depth comparison of computer-mediated and conventional focus group discussions. *International Journal of Market Research*, *47*(2), 131-162.

Reinholt, M., Pedersen, T., & Foss, N.J. (2011). Why a Central Network Position Isn't Enough: The Role of Motivation and Ability for Knowledge Sharing in Employee Networks. *Academy of Management Journal, 54*(6), 1277–1297.

Rodríguez-Garavito, C. (2005). Global Governance and Labor Rights: Codes of Conduct and Anti-sweatshop Struggles in Global Apparel Factories in Mexico and Guatemala. *Politics and Society, 33*(2), 203–233.

Rubin, P.H. (1978). The theory of the firm and the structure of the franchise contract. *Journal of Law & Economics, 21*(1), 223–233.

Rubineau, B., & Fernandez, R.M. (2015). Tipping points: The gender segregating and desegregating effects of network recruitment. *Organization Science, 26*(6), 1646–1664.

Rushe, D. (2016, November 21). Retail group approves Bangladesh factories as safety concerns persist, report finds. *The Guardian (UK)*. https://www.theguardian.com/world/2016/nov/21/bangladesh-garment-factories-safety-alliance-rana-plaza-report, accessed May 2021.

Saxena, S. B. (2018). Beyond Third-party Monitoring: Post-Rana Plaza Interventions. *Economic and Political Weekly, 53*(16), 16-20.

Schneider, H.S. (2012). Agency problems and reputation in expert services: Evidence from auto repair. *Journal of Industrial Economics, 60*(3), 406–433.

Securities and Exchange Commission (2003). Strengthening the commission's requirements regarding auditor independence. 17 CFR 210, 240, 249, and 274 (March 27). https://www.sec.gov/rules/final/33-8183.htm, accessed June 2020.

Seidman, G. (2007). *Beyond the Boycott: Labor Rights, Human Rights, and Transnational Activism.* New York: Russell Sage Foundation.

Seifter, M. (2006). Rent-a-regulator: Design and innovation in privatized governmental decisionmaking. *Ecology Law Quarterly, 33*(4), 1091–1148.

Short, J.L., Toffel, M.W., & Hugill, A.R. (2016). Monitoring global supply chains. *Strategic Management Journal, 37*(9), 1878–1897.

Short, J.L., Toffel, M.W., & Hugill, A.R. (2020). Improving working conditions in global supply chains: The role of institutional environments and monitoring program design. *ILR Review 73*(4), 873–912.

Socher, R., Huval, B., Manning, C.D., & Ng, A.Y. (2012). Semantic compositionality through recursive matrix-vector spaces. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics*, 1201–1211.

Smith, A. (1776). An Inquiry into the Nature and Causes of the Wealth of Nations. London: Printed for W. Strahan, and T. Cadell.

Sun, S., Lu, S.F., & Rui, H. (2020). Does Telemedicine Reduce Emergency Room Congestion? Evidence from New York State. *Information Systems Research, 31*(3), 972–986.

Tan, T.F., & Netessine, S. (2020). At Your Service on the Table: Impact of Tabletop Technology on Restaurant Performance. *Management Science, 66*(10), 4496–4515.

Taylor, F.W. (1911). The Principles of Scientific Management. Harper & brothers.

Tepalagul, N., & Lin, L. (2015). Auditor independence and audit quality: A literature review. *Journal of Accounting, Auditing & Finance, 30*(1), 101–121.

Thomas-Hunt, M.C., Ogden, T.Y., & Neale, M.A. (2003). Who's Really Sharing? Effects of Social and Expert Status on Knowledge Exchange within Groups. *Management Science, 49*(4), 464–477.

Thomson, I. (2007). Mapping the terrain of sustainability accounting. In J. Unerman, J. Bebbington, & B. O'Dwyer (Eds.), *Sustainability, accounting, and accountability* (pp. 19-36). New York: Routledge.

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, and Organization, 2*(2), 181–214.

Toffel, M.W., Short, J.L., & Ouellet, M. (2015). Codes in Context: How States, Markets, and Civil Society Shape Adherence to Global Labor Standards. *Regulation & Governance, 9*(3), 205–223.

Treviño, L.K., Butterfield, K.D., & McCabe, D.L. (1998). The ethical context in organizations: Influences on employee attitudes and behaviors. *Business Ethics Quarterly, 8*(3), 447–476.

Tyre, M.J., & von Hippel, E. (1997). The Situated Nature of Adaptive Learning in Organizations. *Organization Science, 8*(1), 71–83.

U.S. Bureau of Labor Statistics. (2022). U.S. Business Response to the COVID-19 Pandemic – 2021. Report, U.S. Department of Labor, Washington, DC. https://www.bls.gov/news.release/pdf/covid2.pdf

U.S. Department of Health and Human Services. (2022). Conducting Remote Regulatory Assessments. Report, Food and Drug Administration, U.S. Department of Health and Human Services, Washington, DC.

U.S. Department of Labor. (2021). COVID-19: Increased Worksite Complaints and Reduced OSHA Inspections Leave U.S. Workers' Safety at Increased Risk. Report, Office of Inspector General – Office of Audit, U.S. Department of Labor, Washington, DC. https://www.oig.dol.gov/public/reports/oa/2021/19-21-003-10-105.pdf

Utting, P. (2005). Corporate Responsibility and the Movement of Business. *Development in Practice, 15*(3-4), 375–388.

Xue, M., Hitt, L.M., Chen, P. (2011). Determinants and Outcomes of Internet Banking Adoption. *Management Science, 57*(2), 291–307.

Yang, Z., Aydin, G., Babich, V., & Beil, D.R. (2012). Using a dual-sourcing option in the presence of asymmetric information about supplier reliability: Competition vs. diversification. *Manufacturing & Service Operations Management, 14*(2), 202–217.

Supplemental Information for

Essays on Audit Program Design, Audit Quality, and Violation Resolution

A dissertation presented
by
Ashley Kristin Palmarozzo

Ashley Palmarozzo
E-mail: apalmarozzo@hbs.edu

Appendix A – Supplementary Analysis for Chapter 1

Table A-1: Correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13
1 New violations	1.00												
2 Second-party auditor	0.04	1.00											
3 Concurrent source	0.09	0.08	1.00										
4 Different audit firm	0.16	-0.17	0.23	1.00									
5 Different lead auditor (same audit firm)	-0.07	0.04	-0.02	-0.70	1.00								
6 Same lead auditor (same audit firm)	-0.13	0.19	-0.29	-0.49	-0.27	1.00							
7 Female lead auditor	-0.02	-0.09	-0.15	-0.04	0.04	0.00	1.00						
8 Unannounced audit	0.07	0.17	0.07	0.04	-0.05	0.01	-0.11	1.00					
9 Total workers (log)	0.13	0.12	0.05	-0.03	-0.04	0.09	-0.15	0.15	1.00				
10 Limited scope audit	-0.08	0.36	0.00	-0.19	0.08	0.16	0.03	-0.05	0.00	1.00			
11 Re-audit	0.11	-0.31	-0.07	0.10	-0.03	-0.10	0.06	-0.02	-0.08	-0.39	1.00		
12 Follow-up audit	-0.07	0.02	0.12	0.05	-0.01	-0.07	-0.07	0.12	0.08	-0.25	-0.68	1.00	
13 Percent factory supplied to MNC buyer	0.01	0.05	0.03	0.02	-0.04	0.02	-0.07	0.06	-0.11	-0.01	-0.05	0.05	1.00

Table A-2: Summary statistics for samples used in primary tests of H1–H2 (Models 1 and 2 of Table 1.2)

	Model 1 sample to test H1				Model 2 sample to test H2			
	Mean	SD	Min	Max	Mean	SD	Min	Max
New violations	4.95	5.33	0	67	4.43	4.57	0	61
Second-party auditor	0.50	0.50	0	1	0.00	0.00	0	0
Concurrent source	1.00	0.00	1	1	0.90	0.30	0	1
Auditor rotation								
Different audit firm	0.64	0.48	0	1	0.61	0.49	0	1
Different lead auditor (same audit firm)	0.24	0.42	0	1	0.27	0.44	0	1
Same lead auditor (same audit firm)	0.12	0.33	0	1	0.12	0.32	0	1
Female lead auditor	0.40	0.46	0	1	0.41	0.45	0	1
Unannounced audit	0.24	0.43	0	1	0.18	0.39	0	1
Total workers	602	1,099	3	10,601	427	773	2	10,416
Total workers (log)	5.48	1.31	1.4	9.3	5.30	1.18	1.1	9.3
Audit type								
Limited scope audit	0.17	0.38	0	1	0.05	0.22	0	1
Re-audit	0.47	0.50	0	1	0.61	0.49	0	1
Follow-up audit	0.30	0.46	0	1	0.30	0.46	0	1
Percent factory supplied to MNC buyer	0.34	0.24	0.0004	1	0.32	0.25	0.00001	1
Audit sequence	4.38	2.21	2	8	4.07	1.91	2	8
Audit year	2015	1.9	2009	2017	2014	1.8	2007	2017
N (audits)	2,056				7,967			

Table A-3: Covariate balance of matched sample to test H1 in Table 1.2 Model 1

	Total sample				Matched sample			
	Second-party mean	Third-party mean	p-value	Std bias	Second-party mean	Third-party mean	p-value	Std bias
New violations (prior audit)	8.14	5.18	0.00	45.05	5.42	5.45	0.87	-0.74
Total violations (prior audit)	11.29	8.12	0.00	43.95	8.73	8.92	0.45	-3.30
Social audit rating (prior audit)								
Red-Critical rating (prior audit)	0.17	0.04	0.00	46.16	0.04	0.04	1.00	0.02
Red rating (prior audit)	0.52	0.35	0.00	35.91	0.48	0.48	0.98	0.09
Amber rating (prior audit)	0.27	0.54	0.00	-56.69	0.46	0.46	0.98	0.09
Green rating (prior audit)	0.02	0.06	0.00	-23.85	0.01	0.01	1.00	0.01
Concurrent source	0.95	0.90	0.00	19.22	1.00	1.00	1.00	0.00
Different audit firm	0.42	0.61	0.00	-38.66	0.53	0.74	0.00	-44.09
Different lead auditor (same audit firm)	0.31	0.27	0.00	8.83	0.29	0.18	0.00	24.47
Same lead auditor (same audit firm)	0.27	0.12	0.00	39.68	0.18	0.07	0.00	33.31
Second-party auditor (prior audit)	0.58	0.27	0.00	65.65	0.46	0.42	0.04	9.21
Female lead auditor	0.32	0.41	0.00	-19.19	0.39	0.41	0.47	-3.22
Unannounced audit	0.34	0.18	0.00	36.40	0.24	0.24	1.00	0.00
Unannounced audit status missing	*0.05*	*0.01*	*0.00*	*22.96*	*0.01*	*0.01*	*1.00*	*0.00*
Total workers (log)	5.64	5.31	0.00	26.59	5.47	5.49	0.74	-1.44
Total workers	658	430	0.00	24.52	589	614	0.61	-2.27
Audit scope								
Limited scope audit	0.32	0.05	0.00	73.24	0.17	0.17	1.00	0.00
Re-audit	0.27	0.61	0.00	-73.23	0.47	0.47	1.00	0.00
Follow-up audit	0.32	0.30	0.01	5.50	0.30	0.30	1.00	0.00
Missing audit scope	*0.09*	*0.04*	*0.00*	*19.48*	*0.05*	*0.05*	*1.00*	*0.00*
Percent factory supplied to MNC buyer	0.34	0.32	0.00	10.54	0.35	0.33	0.29	4.65
Audit scope (prior audit)								
New factory audit (prior audit)	0.28	0.16	0.00	28.57	0.19	0.19	0.74	1.49
Limited scope audit (prior audit)	0.09	0.08	0.01	5.23	0.07	0.08	0.36	-4.05
Re-audit (prior audit)	0.29	0.41	0.00	-26.74	0.39	0.39	0.82	1.00
Follow-up audit (prior audit)	0.20	0.23	0.00	-6.94	0.23	0.24	0.64	-2.07
Missing audit scope (prior audit)	*0.15*	*0.12*	*0.00*	*6.93*	*0.12*	*0.11*	*0.54*	*2.73*
Audit sequence	3.90	4.06	0.00	-7.81	4.38	4.38	1.00	0.00
Audit year	2014	2014	0.68	0.83	2015	2015	1.00	0.00
N (focal audits)	11,099				2,056			
No. of unbalanced variables	27				4			
Mean standardized bias	9.50				0.46			
Median standardized bias	10.97				0.00			

Notes: p-values obtained from t-tests for continuous variables and from proportion tests for binary variables. Standardized bias is calculated by dividing the mean difference in group means by the average standard deviation. The matched sample was found by exactly matching second-party audits to third-party audits on supplier country, *audit sequence*, *audit year*, *social audit rating (prior audit)*, *concurrent source*, *audit scope,* and *audit announcement* and by coarsened exact matching on *new violations (prior audit).*

Table A-4: Poisson regression results of robustness tests for H1

	(1)	(2)	(3)	(4)
Dependent variable:	New violations	New violations top-coded at 95th percentile	Major new violations	Total violations
Sample:	Full data sample	Matched sample	Matched sample	Matched sample
	Coef	Coef	Coef	Coef
Second-party auditor	0.183**	0.408**	0.434**	0.208**
	(0.032)	(0.062)	(0.072)	(0.037)
Different audit firm	0.254**	0.324**	0.314**	0.092*
	(0.026)	(0.078)	(0.088)	(0.044)
Same lead auditor (same audit firm)	-0.363**	-0.518**	-0.521**	-0.298**
	(0.044)	(0.109)	(0.136)	(0.061)
Female lead auditor	0.160**	0.048	0.039	-0.017
	(0.028)	(0.074)	(0.078)	(0.039)
Unannounced audit	0.113**	0.138	0.238*	0.094
	(0.029)	(0.087)	(0.102)	(0.061)
Total workers (log)	0.067+	0.161+	0.146	0.097+
	(0.036)	(0.092)	(0.106)	(0.056)
Limited scope audit	-0.690**	-0.724**	-0.883**	-0.185+
	(0.060)	(0.197)	(0.242)	(0.112)
Follow-up audit	-0.470**	-0.486**	-0.508**	-0.207**
	(0.028)	(0.070)	(0.082)	(0.041)
Percent factory supplied to MNC buyer	0.023	0.140	0.149	-0.060
	(0.073)	(0.217)	(0.257)	(0.120)
New factory audit (prior audit)	-0.140*	0.176	0.330	0.267*
	(0.064)	(0.210)	(0.261)	(0.127)
Limited scope audit (prior audit)	-0.066	-0.056	-0.046	-0.041
	(0.046)	(0.150)	(0.172)	(0.084)
Follow-up audit (prior audit)	-0.077**	-0.045	-0.086	-0.115*
	(0.027)	(0.074)	(0.086)	(0.045)
Factory fixed effects	Yes	Yes	Yes	Yes
Audit-year fixed effects	Yes	Yes	Yes	Yes
Audit-sequence fixed effects	Yes	Yes	Yes	Yes
N (audits)	10,104	1,241	1,235	1,253
Number of supplier factories	2,321	490	487	496
Sample average of dependent variable	4.6	4.8	4.4	9.7

Notes: Poisson coefficients clustered by supplier factory. + p<.10, * p<.05, **p<.01. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm)*. Omitted category for *new factory audit, limited scope audit*, and *follow-up audit* is *re-audit*.

Table A-5: Poisson regression results based on alternative concurrent sourcing thresholds to test H2

Dependent variable: *new violations*	(1)	(2)	(3)	(4)	(5)	(6)
Thresholds to define concurrent sourcing:	1-99%	5-95%	10-90% (primary approach)	15-85%	20-80%	25-75%
Concurrent source	0.230*	0.206*	0.188**	0.222**	0.204**	0.170**
	(0.111)	(0.081)	(0.073)	(0.075)	(0.045)	(0.061)
Different audit firm	0.299**	0.302**	0.303**	0.297**	0.287**	0.281**
	(0.040)	(0.041)	(0.041)	(0.039)	(0.038)	(0.037)
Same lead auditor (same audit firm)	-0.258**	-0.252**	-0.253**	-0.271**	-0.275**	-0.273**
	(0.072)	(0.074)	(0.074)	(0.075)	(0.074)	(0.072)
Female lead auditor	0.125**	0.127**	0.125**	0.128**	0.125**	0.121**
	(0.045)	(0.045)	(0.046)	(0.043)	(0.044)	(0.046)
Unannounced audit	0.145**	0.148**	0.150**	0.150**	0.145**	0.145**
	(0.052)	(0.051)	(0.050)	(0.051)	(0.047)	(0.046)
Total workers (log)	0.067	0.066	0.067	0.073+	0.072+	0.078+
	(0.044)	(0.044)	(0.044)	(0.044)	(0.044)	(0.046)
Limited scope audit	-0.631**	-0.628**	-0.623**	-0.610**	-0.608**	-0.605**
	(0.096)	(0.095)	(0.093)	(0.086)	(0.085)	(0.094)
Follow-up audit	-0.464**	-0.462**	-0.462**	-0.462**	-0.467**	-0.461**
	(0.043)	(0.043)	(0.043)	(0.043)	(0.042)	(0.043)
Percent factory supplied to MNC buyer	0.035	0.039	0.036	0.040	0.053	0.049
	(0.075)	(0.075)	(0.074)	(0.077)	(0.078)	(0.074)
New factory audit (prior audit)	-0.077	-0.077	-0.074	-0.076	-0.072	-0.063
	(0.052)	(0.052)	(0.052)	(0.053)	(0.054)	(0.057)
Limited scope audit (prior audit)	-0.058	-0.060	-0.060	-0.055	-0.046	-0.038
	(0.041)	(0.041)	(0.042)	(0.042)	(0.043)	(0.044)
Follow-up audit (prior audit)	-0.127**	-0.126**	-0.125**	-0.126**	-0.129**	-0.123*
	(0.047)	(0.048)	(0.047)	(0.047)	(0.047)	(0.049)
Factory fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Audit-year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Audit-sequence fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
N (audits)	7,009	7,009	7,009	7,009	7,009	7,009
Supplier factories	1,847	1,847	1,847	1,847	1,847	1,847
Sample average of dependent variable	4.295	4.295	4.295	4.295	4.295	4.295

Notes: Poisson coefficients with robust standard errors clustered at the supplier-factory country. + $p<0.10$, * $p<0.05$, ** $p<0.01$. To isolate the effect of transitioning from a third-party market to a concurrent sourcing market, we exclude from each sample the few instances in which the reverse transition occurred. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm)*. Omitted category for *new factory audit*, *limited scope audit*, and *follow-up audit* is *re-audit*.

Table A-6: Poisson regression results of robustness tests for H2

	(1)	(2)	(3)
Dependent variable:	New violations top-coded at 95th percentile	Major new violations	Total violations
Concurrent source	0.200**	0.232*	0.156**
	(0.072)	(0.097)	(0.038)
Different audit firm	0.291**	0.316**	0.131**
	(0.034)	(0.041)	(0.034)
Same lead auditor (same audit firm)	-0.238**	-0.240**	-0.051
	(0.073)	(0.081)	(0.045)
Female lead auditor	0.132**	0.135*	0.041
	(0.042)	(0.067)	(0.044)
Unannounced audit	0.147**	0.171**	0.099**
	(0.049)	(0.047)	(0.035)
Total workers (log)	0.072+	0.059	0.008
	(0.038)	(0.048)	(0.035)
Limited scope audit	-0.561**	-0.646**	-0.219**
	(0.063)	(0.107)	(0.067)
Follow-up audit	-0.450**	-0.484**	-0.134**
	(0.043)	(0.038)	(0.014)
Percent factory supplied to MNC buyer	0.032	0.008	0.020
	(0.066)	(0.054)	(0.056)
New factory audit (prior audit)	-0.085+	-0.048	0.025
	(0.048)	(0.061)	(0.053)
Limited scope audit (prior audit)	-0.053	-0.091+	-0.090**
	(0.033)	(0.050)	(0.033)
Follow-up audit (prior audit)	-0.107**	-0.143**	-0.093**
	(0.036)	(0.052)	(0.021)
Factory fixed effects	Yes	Yes	Yes
Audit-year fixed effects	Yes	Yes	Yes
Audit-sequence fixed effects	Yes	Yes	Yes
N (audits)	7,009	6,941	7,037
Supplier factories	1,847	1,822	1,859
Sample average of dependent variable	4.2	3.7	8.5

Notes: Poisson coefficients with robust standard errors clustered at the supplier-factory country. + p<0.10, * p<0.05, **p<0.01. To isolate the effect of transitioning from a third-party market to a concurrent sourcing market, we exclude from each sample the few instances in which the reverse transition occurred. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm)*. Omitted category for *new factory audit*, *limited scope audit*, and *follow-up audit* is *re-audit*.

Table A-7: Poisson regression results of robustness tests for H3

	(1)	(2)	(3)	(4)	(5)
Dependent variable:	New violations	New violations	New violations top-coded at 95th percentile	Major new violations	Total violations
Sample:	See notes	Matched sample	All audits	All audits	All audits
Second-party auditor	0.171**	0.250**	0.165**	0.178**	0.109**
	(0.034)	(0.070)	(0.029)	(0.035)	(0.019)
Concurrent source	0.178*	0.029	0.150*	0.139	0.054
	(0.087)	(0.192)	(0.072)	(0.098)	(0.053)
Different audit firm	0.243**	0.297**	0.248**	0.274**	0.104**
	(0.029)	(0.043)	(0.024)	(0.028)	(0.015)
Same lead auditor (same audit firm)	-0.377**	-0.279**	-0.331**	-0.353**	-0.114**
	(0.047)	(0.090)	(0.038)	(0.049)	(0.027)
Female lead auditor	0.180**	0.111*	0.153**	0.139**	0.041**
	(0.031)	(0.048)	(0.026)	(0.029)	(0.015)
Unannounced audit	0.136**	0.204*	0.101**	0.149**	0.095**
	(0.033)	(0.079)	(0.027)	(0.031)	(0.018)
Total workers (log)	0.066+	0.086	0.059+	0.054	0.013
	(0.038)	(0.076)	(0.032)	(0.038)	(0.021)
Limited scope audit	-0.705**	-0.731**	-0.623**	-0.733**	-0.279**
	(0.061)	(0.145)	(0.052)	(0.064)	(0.040)
Follow-up audit	-0.496**	-0.455**	-0.456**	-0.492**	-0.142**
	(0.031)	(0.050)	(0.026)	(0.030)	(0.016)
Percent factory supplied to MNC buyer	0.004	0.041	0.037	-0.007	0.002
	(0.080)	(0.156)	(0.069)	(0.079)	(0.048)
New factory audit (prior audit)	-0.124+	-0.127	-0.153*	-0.127+	-0.041
	(0.068)	(0.137)	(0.060)	(0.070)	(0.042)
Limited scope audit (prior audit)	-0.035	-0.044	-0.068	-0.104*	-0.100**
	(0.048)	(0.118)	(0.043)	(0.049)	(0.031)
Follow-up audit (prior audit)	-0.064*	-0.080	-0.061*	-0.098**	-0.077**
	(0.032)	(0.051)	(0.025)	(0.029)	(0.016)
Factory fixed effects	Yes	Yes	Yes	Yes	Yes
Audit-year fixed effects	Yes	Yes	Yes	Yes	Yes
Audit-sequence fixed effects	Yes	Yes	Yes	Yes	Yes
N (audits)	8,826	2,949	10,104	10,045	10,133
Number of supplier factories	2,321	943	2,321	2,300	2,335
Sample average of dependent variable	4.5	4.1	4.4	3.9	9.2

Notes: The sample of the model reported in Column 1 excludes audits in which the factory's improvement score, calculated as the percent reduction in number of last-audit *new violations* from the average *new violations* recorded in a factory's preceding two audits, is within the largest 25th percentile of improvement scores.

Poisson coefficients with robust standard errors clustered at the supplier. + p<0.10, * p<0.05, **p<0.01. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm)*. Omitted category for audit type *new factory audit, limited scope audit*, and *follow-up audit* is *re-audit*.

Table A-8: Poisson regression results to assess the impact of familiarity on audit firm rotation

Dependent variable = *new violations*	(1)	
	Coef	AME
New different audit firm	0.279**	1.3
	(0.027)	
Familiar different audit firm	0.154**	0.7
	(0.035)	
Same lead auditor (same audit firm)	-0.362**	-1.5
	(0.044)	
Second-party auditor	0.190**	0.9
	(0.032)	
Concurrent source	0.139+	0.6
	(0.083)	
Female lead auditor	0.159**	0.7
	(0.028)	
Unannounced audit	0.110**	0.5
	(0.029)	
Total workers (log)	0.069+	0.3
	(0.036)	
Limited scope audit	-0.691**	-2.5
	(0.060)	
Follow-up audit	-0.468**	-2.0
	(0.028)	
Percent factory supplied to MNC buyer	0.025	0.1
	(0.073)	
New factory audit (prior audit)	-0.138*	-0.6
	(0.064)	
Limited scope audit (prior audit)	-0.069	-0.3
	(0.046)	
Follow-up audit (prior audit)	-0.072**	-0.3
	(0.027)	
Factory fixed effects	Yes	
Audit-year fixed effects	Yes	
Audit-sequence fixed effects	Yes	
N (audits)	10,104	
Supplier factories	2,321	
Sample average of dependent variable	4.6	

Notes: Poisson coefficients with robust standard errors clustered at the supplier. + p<0.10, * p<0.05, **p<0.01. Omitted category for *different audit firm* and *same lead auditor (same audit firm)* is *different lead auditor (same audit firm)*. Omitted category for *new factory audit, limited scope audit*, and *follow-up audit* is *re-audit*. AME is average marginal effects.

Figure A-9: Kernel density graphs illustrating balance in the primary matched sample used to test H1

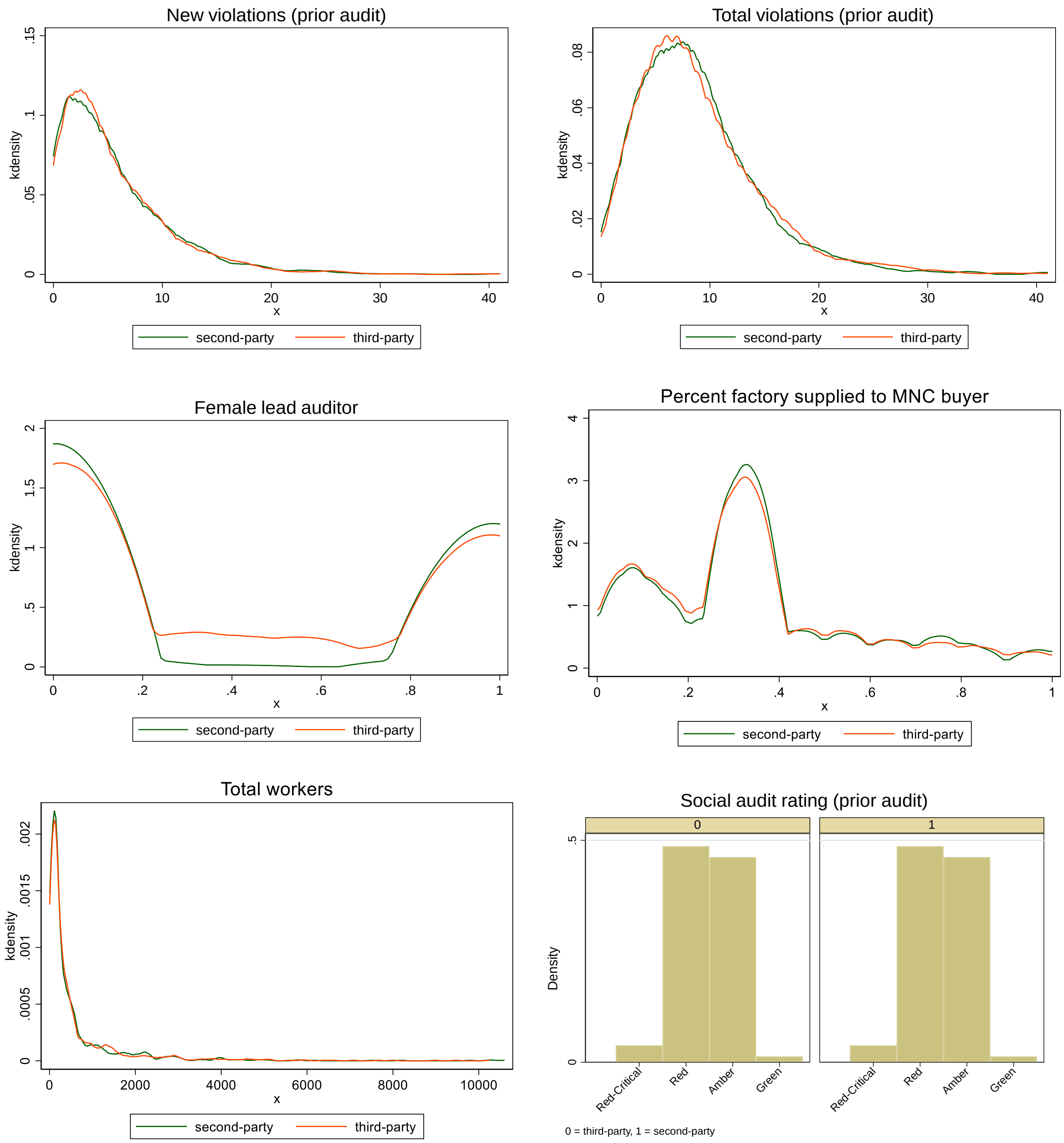
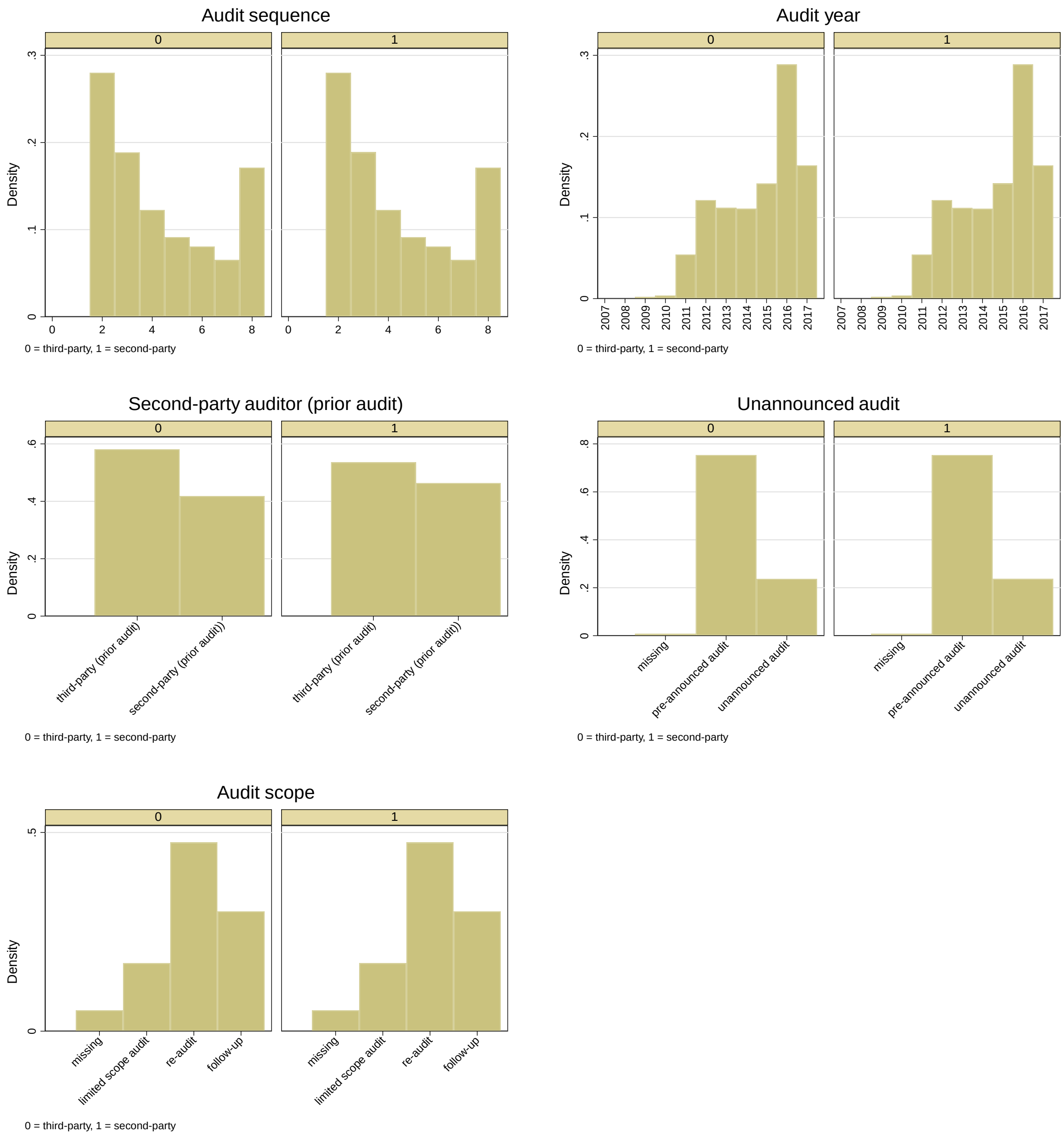


Figure A-9: Kernel density graphs illustrating balance in the primary matched sample used to test H1 (continued)



Appendix B – Supplementary Information on Sample Construction for Chapter 2

Table B-5: Sample Construction

	Notes	Audits	% loss in data from prior row
All 2019-2021 surveillance audits for the six standards	[1]	125,727	
Omit if the surveillance audit is a site's only one for that standard	[2]	101,419	19.3%
Omit data anomalies:			
Omit surveillance audits missing the certificate number	[3]	101,317	0.1%
Omit duplicate surveillance audits (same audit format-site-standard-date)	[4]	99,697	1.6%
Omit surveillance audits that prompt a change in certificate number.	[5]	98,290	1.4%
Omit surveillance audits from site-standards with consecutive "initial stage 1" audits and consecutive "initial stage 2" audits	[6]	97,380	0.9%
Omit surveillance audits with an incorrect audit sequence or is missing a step missing in the audit sequence	[7]	96,231	0.1%
Omit surveillance audits where the auditor ID is missing	[8]	96,222	0.0%
Omit surveillance audits if the number of surveillance audits per cycle exceeds two, or if the number of surveillance audits per year exceeds one	[9]	67,814	33.9%
Omit if all of a site's surveillance audits of a particular standard yield zero violations	[10]	35,247	50.2%

Notes:

[1]: Standards include ISO 9001, ISO 14001, ISO 27001, OHSAS 18001/ISO 45001, and ISO 13485.

[2]: Because our model specification compares within-site audits of the same standard (by using site-standard fixed effects), our model requires two audits per site-standard.

[3]: Audits linked to a specific site and standard are identified by their having identical standard certificate numbers. Without a certificate number, we cannot identify whether audits observed for the same site and standard are part of the same, or different, certificates, which prevents us from identifying other potential anomalies within the certificate, such as the data anomalies in [4]-[9].

[4]: This reflects audits on the same date, for the same site, and for the same standard that are both remote or in-person. These audits may reflect two partial audits or one entire audit mistakenly recorded twice.

[5]: Certificate numbers typically change after a re-certification audit. Changes to the certificate number following a surveillance audit could reflect major operational changes in the audited site, which could affect the number of violations and thus confound our analysis.

[6]: Site-standards with initial stage 1 or 2 audits could occur for a few reasons: the site required more than one of that audit type to "pass" it or the audit type was incorrectly coded. Because we can't distinguish whether the consecutive initial audit is a mis-coded surveillance audit, we drop surveillance audits from site-standards that display this anomaly.

[7]: Audit sequences should include an initial stage 1 audit to assess a site's readiness for certification, an initial stage 2 audit to recommend an establishment for certification, two surveillance audits, then a re-certification audit (subsequently, the sequence becomes two surveillance audits followed by a re-certification audit). Scenarios where we found an incorrect audit sequence include: an initial stage 2 audit occurring before an

Table B-6: Sample Construction (continued)

initial stage 1 audit, a surveillance audit occurring before an initial stage 1 or 2 audit, or a re-certification audit occurring before an initial stage 1 or 2 audit. Scenarios where we found missing steps in the audit sequence include: an initial stage 1 audit followed directly by a surveillance audit (missing the initial stage 2 audit), an initial stage 1 audit followed by a recertification (missing the initial stage 2 audit and both surveillance audits), and an initial stage 2 audit followed directly by a re-certification audit (missing both surveillance audits).

[8]: Observing the auditor ID is crucial to our linking to auditor demographic data and is crucial to developing several key variables, such as audit team size, auditor qualifications, audit team gender makeup, an auditor's prior experience with the audited site, and cumulative auditing experience.

[9]: We omit site-standards where we observe more than the expected two surveillance audits between (a) an initial audit and a re-certification audit, or (b) two re-certification audits. When there are more than two consecutive surveillance audits, some of them are likely partial audits with more limited scope, which can (mechanically) reduce violations and confound our analysis.

[10]: Poisson regression with site-standard fixed effects, which we use to estimate our models, drops all audits pertaining to a site-standard dyad if they all report zero violations.

Appendix C – Extension Results for Chapter 2

Table C-1: Extension Regression Results

	(1)	(2)
remote audit	-0.292**	-0.377**
	(0.026)	(0.042)
remote audit X management standard focus (log)		0.145**
		(0.055)
maximum management standard focus	0.113*	0.056
	(0.055)	(0.059)
multi-member audit team	0.028	0.030
	(0.041)	(0.041)
average prior in-person site exposure (log)	-0.135**	-0.136**
	(0.021)	(0.021)
focal standard advanced training[a]	0.035	0.035
	(0.034)	(0.034)
staff-days (log)	0.420**	0.422**
	(0.051)	(0.051)
COVID time period	-0.065+	-0.066+
	(0.039)	(0.039)
prior remote site exposure	-0.038	-0.030
	(0.029)	(0.029)
maximum auditing experience (log)	-0.030*	-0.030*
	(0.012)	(0.013)
female on audit team[a]	0.078*	0.077*
	(0.035)	(0.035)
percent outsourced	-0.115**	-0.113**
	(0.040)	(0.040)
multi-standard audit	-0.021	-0.021
	(0.041)	(0.041)
Site-standard fixed effects	Included	Included
Audit year fixed effects	Included	Included
Audit sequence fixed effects	Included	Included
Number of audits	35,247	35,247
Number of audited sites	14,615	14,615
Number of audited site-standards	16,986	16,986

Notes: The unit of analysis is audited site-standard-audit. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Appendix D – Robustness Results for Chapter 2

Table D-1: Regression results, omission of remote-to-in-person audit sequence

	Dependent variable: Violations	(1)	(2)	(3)	(4)
H1	remote audit	-0.264**	-0.200**	-0.210**	-0.370**
		(0.029)	(0.032)	(0.030)	(0.042)
H2	remote audit X observation		-0.400**		
			(0.046)		
H3	remote audit X multi-member audit team			-0.307**	
				(0.047)	
H4	remote audit X average prior in-person site exposure (log)				0.093**
					(0.027)
	multi-member audit team	0.033	0.068	0.133**	0.028
		(0.042)	(0.048)	(0.046)	(0.042)
	average prior in-person site exposure (log)	-0.135**	-0.126**	-0.136**	-0.198**
		(0.021)	(0.023)	(0.021)	(0.029)
	focal standard advanced training[a]	0.040	0.025	0.042	0.040
		(0.035)	(0.041)	(0.035)	(0.035)
	staff-days (log)	0.428**	0.441**	0.430**	0.428**
		(0.053)	(0.062)	(0.054)	(0.052)
	COVID time period	-0.070+	-0.248**	-0.078*	-0.073+
		(0.039)	(0.042)	(0.039)	(0.039)
	prior remote site exposure	-0.047	-0.057+	-0.040	-0.063*
		(0.030)	(0.033)	(0.030)	(0.030)
	maximum auditing experience (log)	-0.036**	-0.045**	-0.036**	-0.032**
		(0.012)	(0.014)	(0.012)	(0.013)
	female on audit team[a]	0.068+	0.070+	0.074*	0.067+
		(0.036)	(0.039)	(0.036)	(0.036)
	percent outsourced	-0.107**	-0.112*	-0.104**	-0.111**
		(0.040)	(0.045)	(0.040)	(0.040)
	multi-standard audit	-0.033	-0.036	-0.023	-0.038
		(0.042)	(0.047)	(0.043)	(0.042)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	33,573	35,849	33,573	33,573
	Number of audited sites	14,008	12,749	14,008	14,008
	Number of audited site-standards	16,244	14,718	16,244	16,244

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3, and 4, and is at the audited site-standard-detection mode for column 2. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-2: Regression results, violations from mandatory clauses

Dependent variable: Violations from Mandatory Clauses		(1)	(2)	(3)	(4)
H1	remote audit	-0.346**	-0.254**	-0.283**	-0.441**
		(0.035)	(0.037)	(0.036)	(0.053)
H2	remote audit X observation		-0.506**		
			(0.052)		
H3	remote audit X multi-member audit team			-0.388**	
				(0.063)	
H4	remote audit X average prior in-person site exposure (log)				0.082*
					(0.035)
	multi-member audit team	-0.010	-0.007	0.113+	-0.013
		(0.057)	(0.058)	(0.061)	(0.057)
	average prior in-person site exposure (log)	-0.104**	-0.099**	-0.106**	-0.160**
		(0.027)	(0.028)	(0.027)	(0.037)
	focal standard advanced training[a]	0.049	0.038	0.052	0.047
		(0.050)	(0.051)	(0.051)	(0.050)
	staff-days (log)	0.581**	0.584**	0.581**	0.580**
		(0.068)	(0.071)	(0.068)	(0.068)
	COVID time period	-0.204**	-0.231**	-0.212**	-0.207**
		(0.051)	(0.052)	(0.051)	(0.051)
	prior remote site exposure	-0.005	-0.020	0.002	-0.016
		(0.039)	(0.040)	(0.039)	(0.039)
	maximum auditing experience (log)	-0.043**	-0.049**	-0.042**	-0.039*
		(0.016)	(0.017)	(0.016)	(0.016)
	female on audit team[a]	0.086+	0.099*	0.085+	0.086+
		(0.045)	(0.045)	(0.045)	(0.045)
	percent outsourced	-0.086+	-0.091+	-0.085	-0.090+
		(0.052)	(0.053)	(0.052)	(0.052)
	multi-standard audit	-0.015	-0.019	-0.001	-0.017
		(0.059)	(0.061)	(0.059)	(0.059)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	24,581	27,622	24,581	24,581
	Number of audited sites	10,615	10,412	10,615	10,615
	Number of audited site-standards	11,834	11,609	11,834	11,834

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3, and 4, and is at the audited site-standard-detection mode for column 2. The dependent variable reflects the number of audit violations recorded from audit standard clauses that must be checked in each surveillance audit. As such, the N for all models is smaller than the N in our main results (Table 2.3) because our omitting violations from non-mandatory standard clauses increases the number of site-standards that observe no mandatory violations for all audits in the sample, and audits for these site-standards will be omitted from the regression due to our fixed effects specification. The N in model 2 is less than twice as large the N in all other models because we omit ISO 13485 audits (we cannot map violations to the primary detection mode for this standard) and because our inclusion of the three-way fixed effect (as opposed to the two-way fixed effect included in all other models) omits additional observations when there are no violations recorded within a site, standard, and detection mode. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-3: Regression results, minor violations

Dependent variable: Minor Violations		(1)	(2)	(3)	(4)
H1	remote audit	-0.294**	-0.227**	-0.246**	-0.396**
		(0.026)	(0.029)	(0.027)	(0.039)
H2	remote audit X observation		-0.425**		
			(0.045)		
H3	remote audit X multi-member audit team			-0.388**	
				(0.063)	
H4	remote audit X average prior in-person site exposure (log)				0.082*
					(0.035)
	multi-member audit team	-0.010	0.072	0.113+	-0.013
		(0.057)	(0.047)	(0.061)	(0.057)
	average prior in-person site exposure (log)	-0.128**	-0.127**	-0.130**	-0.190**
		(0.020)	(0.023)	(0.020)	(0.027)
	focal standard advanced training[a]	0.040	0.024	0.042	0.041
		(0.034)	(0.040)	(0.035)	(0.034)
	staff-days (log)	0.434**	0.432**	0.437**	0.435**
		(0.051)	(0.060)	(0.052)	(0.051)
	COVID time period	-0.066+	-0.241**	-0.073+	-0.069+
		(0.039)	(0.042)	(0.039)	(0.039)
	prior remote site exposure	-0.028	-0.044	-0.021	-0.042
		(0.029)	(0.032)	(0.029)	(0.029)
	maximum auditing experience (log)	-0.033**	-0.044**	-0.033**	-0.029*
		(0.012)	(0.014)	(0.012)	(0.012)
	female on audit team[a]	0.075*	0.075*	0.080*	0.074*
		(0.034)	(0.038)	(0.034)	(0.034)
	percent outsourced	-0.115**	-0.119**	-0.113**	-0.118**
		(0.040)	(0.044)	(0.040)	(0.040)
	multi-standard audit	-0.017	-0.023	-0.005	-0.022
		(0.041)	(0.046)	(0.041)	(0.041)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	35,172	37,478	35,172	35,172
	Number of audited sites	14,586	13,263	14,586	14,586
	Number of audited site-standards	16,949	15,345	16,949	16,949

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3, and 4, and is at the audited site-standard-detection mode for column 2. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-4: Regression results, include bundled audits

Dependent variable: Violations		(1)	(2)	(3)	(4)
H1	remote audit	-0.328**	-0.282**	-0.294**	-0.375**
		(0.023)	(0.026)	(0.023)	(0.034)
H2	remote audit X observation		-0.344**		
			(0.037)		
H3	remote audit X multi-member audit team			-0.388**	
				(0.063)	
H4	remote audit X average prior in-person site exposure (log)				0.082*
					(0.035)
	multi-member audit team	-0.010	0.078+	0.113+	-0.013
		(0.057)	(0.041)	(0.061)	(0.057)
	average prior in-person site exposure (log)	-0.140**	-0.131**	-0.140**	-0.167**
		(0.017)	(0.019)	(0.017)	(0.023)
	focal standard advanced training[a]	0.045	0.041	0.047+	0.045
		(0.028)	(0.033)	(0.028)	(0.028)
	staff-days (log)	0.432**	0.386**	0.431**	0.433**
		(0.046)	(0.053)	(0.047)	(0.046)
	COVID time period	0.016	-0.087**	0.009	0.016
		(0.029)	(0.032)	(0.029)	(0.029)
	prior remote site exposure	-0.053*	-0.065**	-0.050*	-0.060**
		(0.023)	(0.025)	(0.023)	(0.023)
	maximum auditing experience (log)	-0.030**	-0.040**	-0.031**	-0.028*
		(0.011)	(0.013)	(0.011)	(0.011)
	female on audit team[a]	0.094**	0.097**	0.097**	0.093**
		(0.028)	(0.031)	(0.028)	(0.028)
	percent outsourced	-0.103**	-0.098**	-0.099**	-0.106**
		(0.033)	(0.037)	(0.033)	(0.033)
	multi-standard audit	0.002	0.007	0.007	0.000
		(0.035)	(0.037)	(0.035)	(0.035)
	hybrid audit	0.193**	0.199**	0.186**	0.182**
		(0.033)	(0.036)	(0.033)	(0.033)
	bundled audit	0.173**	0.283**	0.174**	0.174**
		(0.048)	(0.054)	(0.048)	(0.048)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	43,220	47,484	43,220	43,220
	Number of audited sites	16,988	15,634	16,988	16,988
	Number of audited site-standards	20,538	18,673	20,538	20,538

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3, and 4, and is at the audited site-standard-detection mode for column 2. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-5: Regression results, use average version of variable

Dependent variable: Violations		(1)	(2)	(3)	(4)
H1	remote audit	-0.293**	-0.228**	-0.244**	-0.392**
		(0.026)	(0.029)	(0.027)	(0.040)
H2	remote audit X observation		-0.425**		
			(0.045)		
H3	remote audit X multi-member audit team			-0.388**	
				(0.063)	
H4	remote audit X average prior in-person site exposure (log)				0.082*
					(0.035)
	multi-member audit team	0.023	0.057	0.122**	0.020
		(0.041)	(0.047)	(0.044)	(0.041)
	average prior in-person site exposure (log)	-0.136**	-0.128**	-0.138**	-0.196**
		(0.021)	(0.023)	(0.021)	(0.028)
	focal standard advanced training[a]	0.042	0.026	0.043	0.042
		(0.034)	(0.040)	(0.034)	(0.034)
	staff-days (log)	0.421**	0.431**	0.423**	0.422**
		(0.051)	(0.060)	(0.052)	(0.051)
	COVID time period	-0.065+	-0.241**	-0.072+	-0.068+
		(0.039)	(0.042)	(0.039)	(0.039)
	prior remote site exposure	-0.038	-0.045	-0.031	-0.051+
		(0.029)	(0.032)	(0.029)	(0.029)
	average auditing experience (log)	-0.030*	-0.042**	-0.031*	-0.027*
		(0.012)	(0.014)	(0.012)	(0.012)
	female on audit team[a]	0.082*	0.075*	0.087*	0.081*
		(0.035)	(0.038)	(0.035)	(0.035)
	percent outsourced	-0.109**	-0.119**	-0.107**	-0.112**
		(0.039)	(0.044)	(0.040)	(0.040)
	multi-standard audit	-0.022	-0.024	-0.011	-0.027
		(0.041)	(0.046)	(0.042)	(0.041)
	percent outsourced	-0.109**	-0.119**	-0.107**	-0.112**
		(0.039)	(0.044)	(0.040)	(0.040)
	multi-standard audit	-0.022	-0.024	-0.011	-0.027
		(0.041)	(0.046)	(0.042)	(0.041)
	Audited site-standard fixed effects	Included		Included	Included
	Audited site-standard- detection mode fixed effects		Included		
	Audit year fixed effects	Included	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included	Included
	Number of audits	35,247	37,631	35,247	35,247
	Number of audited sites	14,615	13,296	14,615	14,615
	Number of audited site-standards	16,986	15,388	16,986	16,986

Notes: The unit of analysis is at the audited site-standard-audit for columns 1, 3, and 4, and is at the audited site-standard-detection mode for column 2. The N in model 2 is less than twice as large the N in all other models because we omit ISO 13485 audits (we cannot map violations to the primary detection mode for this standard) and because our inclusion of the three-way fixed effect (as opposed to the two-way fixed effect included in all other models) omits additional observations when there are no violations recorded within a site, standard, and detection mode. In model 2, the omitted category for detection mode is *document review* (and thus provides the baseline comparison for *observation*). In model 2, the main effect of *observation* is absorbed by the inclusion of site-standard-detection mode fixed effects. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-6: Regression results, omit ISO 13485 Audits, hypotheses 1, 3-4

Dependent variable: Violations		(1)	(2)	(3)
H1	remote audit	-0.288**	-0.236**	-0.391**
		(0.027)	(0.027)	(0.041)
H3	remote audit X multi-member audit team		-0.320**	
			(0.048)	
H4	remote audit X average prior in-person site exposure (log)			0.089**
				(0.026)
	multi-member audit team	0.065	0.171**	0.062
		(0.044)	(0.048)	(0.044)
	average prior in-person audit exposure (log)	-0.130**	-0.132**	-0.191**
		(0.021)	(0.022)	(0.029)
	focal standard advanced training[a]	0.019	0.022	0.017
		(0.037)	(0.038)	(0.037)
	staff-days (log)	0.435**	0.439**	0.436**
		(0.055)	(0.056)	(0.054)
	COVID time period	-0.080*	-0.087*	-0.083*
		(0.040)	(0.040)	(0.040)
	prior remote site exposure	-0.031	-0.023	-0.045
		(0.030)	(0.030)	(0.030)
	maximum auditing experience (log)	-0.039**	-0.039**	-0.035**
		(0.013)	(0.013)	(0.013)
	female on audit team[a]	0.074*	0.077*	0.073*
		(0.036)	(0.036)	(0.036)
	percent outsourced	-0.107**	-0.105*	-0.111**
		(0.041)	(0.041)	(0.041)
	multi-standard audit	-0.015	-0.003	-0.020
		(0.042)	(0.042)	(0.042)
	Audited site-standard fixed effects	Included	Included	Included
	Audit year fixed effects	Included	Included	Included
	Audit sequence fixed effects	Included	Included	Included
	Number of audits	33,758	33,758	33,758
	Number of audited sites	13,915	13,915	13,915
	Number of audited site-standards	16,244	16,244	16,244

Notes: The unit of analysis is at the audited site-standard. The N in this table is less than that in the main results (Table 2.3) because in this sample we omit audits from ISO 13485. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **p<0.01; *p<0.05; +p<0.10.

[a] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Table D-7: Regression results, top-code variable at 95th percentile (H4)

Dependent variable: Violations		(1)	(2)
H1	remote audit	-0.295**	-0.405**
		(0.026)	(0.040)
H4	remote audit X average prior in-person site exposure (log)		0.098**
			(0.027)
	multi-member audit team	0.036	0.032
		(0.041)	(0.041)
	average prior in-person site exposure (log)[a]	-0.143**	-0.210**
		(0.021)	(0.028)
	focal standard advanced training[b]	0.041	0.042
		(0.034)	(0.034)
	staff-days (log)	0.422**	0.423**
		(0.051)	(0.051)
	COVID time period	-0.065+	-0.069+
		(0.039)	(0.039)
	prior remote site exposure	-0.038	-0.051+
		(0.029)	(0.029)
	maximum auditing experience (log)	-0.032**	-0.028*
		(0.012)	(0.012)
	female on audit team[b]	0.082*	0.080*
		(0.035)	(0.035)
	percent outsourced	-0.109**	-0.112**
		(0.039)	(0.040)
	multi-standard audit	-0.020	-0.025
		(0.041)	(0.041)
	Audited site-standard fixed effects	Included	Included
	Audited site-standard- detection mode fixed effects		
	Audit year fixed effects	Included	Included
	Audit sequence fixed effects	Included	Included
	Number of audits	35,247	35,247
	Number of audited sites	14,615	14,615
	Number of audited site-standards	16,986	16,986

Notes: The unit of analysis is at the audited site-standard-audit. All models are estimated using Poisson regression. Standard errors are clustered by audited site. **$p<0.01$; *$p<0.05$; +$p<0.10$.

[a] denotes variables that are top-coded at the 95th percentile.

[b] denotes variables where missing values are recoded to zero; all models include a series of dummy variables that indicates the observations where such recoding occurred for each variable.

Appendix E – Chapter 3 Robustness Test Results

Table E-1: Robustness results, major violations

DEPENDENT VARIABLE= Resolved	(1)	OR
percent supplied (prior audit)[1]	0.201***	1.22
	(0.075)	
number of alternative suppliers (prior audit)[2]	0.025***	1.03
	(0.009)	
second-party (prior audit)	-0.071	0.93
	(0.045)	
second-party	-0.128***	0.88
	(0.049)	
unannounced (prior audit)	0.013	1.01
	(0.058)	
unannounced	-0.233***	0.79
	(0.056)	
number of recorded violations (prior audit)	-0.015***	0.99
	(0.003)	
new factory audit (prior audit) [1]	-0.060	0.94
	(0.066)	
follow-up audit (prior audit) [1]	0.053	1.05
	(0.046)	
follow-up audit[1]	0.079	1.08
	(0.061)	
female lead auditor (prior audit) [1]	0.014	1.01
	(0.041)	
female lead auditor[1]	-0.051	0.95
	(0.045)	
first time lead auditor (prior audit)	0.130**	1.14
	(0.056)	
first time lead auditor	0.051	1.05
	(0.053)	
total workers log (prior audit) [1,2]	0.056***	1.06
	(0.017)	
percent female workers (prior audit) [1]	0.140	1.15
	(0.138)	
factory tenure (prior audit) [1]	-0.008	0.99
	(0.007)	
audit sequence	0.064***	1.07
	(0.016)	
FIXED EFFECTS		
violation vintage	Included	
violation functional sub-category	Included	

Table E-1: Robustness results, major violations (continued)

Country	Included
industry	included
audit year	Included
RANDOM EFFECTS	
audit	Included
factory	Included

Notes: *p<0.10, **p<0.05, ***p<0.01; N = 62,056 observations. The number of observations in this model is less than that of the main model results because this sample excludes minor violations. The coefficient for *major* is therefore omitted from this model given that the sample reflects only major violations. The model is estimated using a multilevel mixed-effects logistic regression model with standard errors clustered at the factory. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model.

Table E-2: Robustness results, modeling factory fixed effects

	(1)	OR
percent supplied (prior audit)[1]	0.293**	1.34
	(0.120)	
number of alternative suppliers (prior audit)	0.048***	1.05
	(0.017)	
second-party (prior audit)	-0.109**	0.90
	(0.050)	
second-party	-0.083*	0.92
	(0.049)	
unannounced (prior audit)	-0.134*	0.87
	(0.069)	
unannounced	-0.162***	0.85
	(0.053)	
number of recorded violations (prior audit)	0.021***	1.02
	(0.004)	
new factory audit (prior audit) [1]	0.147**	1.16
	(0.067)	
follow-up audit (prior audit) [1]	0.246***	1.28
	(0.046)	
follow-up audit[1]	0.030	1.03
	(0.060)	
female lead auditor (prior audit) [1]	-0.060	0.94
	(0.046)	
female lead auditor[1]	-0.038	0.96
	(0.045)	
first time lead auditor (prior audit)	0.005	1.01
	(0.055)	
first time lead auditor	0.053	1.05
	(0.053)	
total workers log (prior audit) [1]	0.036	1.04
	(0.029)	
percent female workers (prior audit) [1]	0.518***	1.68
	(0.168)	
factory tenure (prior audit) [1]	-0.022	0.98
	(0.034)	
Major violation	0.308	1.36
	(0.269)	
audit sequence	0.096***	1.10
	(0.034)	
FIXED EFFECTS		
violation vintage	Included	

Table E-2: Robustness results, modeling factory fixed effects (continued)

violation severity-category	Included
audit year	Included
factory	Included

Notes: *p<0.10, **p<0.05, ***p<0.01; N = 70,488 observations from 8,179 factory audits of 2,282 factories. This model is estimated using a conditional logistic regression model with factory fixed effects and standard errors clustered at the factory. The number of observations in this model is less than that of the main model results due to changes to model specification: this model specification include factory fixed effects which omits 437 audits of 344 factories where no violations are resolved before the end of the data panel. Including all of the independent and control variables in a conditional logistic regression that also included 2,282 factory fixed effects variables led to non-convergence in the model estimation, so I omitted the 92 violation functional sub-category fixed effects and instead included 24 dummy variables that captured all combinations of violation severity (major or minor) and violation functional parent category. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model.

Table E-3: Robustness results, alternative H2 measure

DEPENDENT VARIABLE= Resolved	(1)	OR
percent supplied (prior audit)[1]	0.195***	1.22
	(0.072)	
number of alternative suppliers in prior two years (prior audit)'[2]	0.017**	1.02
	(0.008)	
second-party (prior audit)	-0.095**	0.91
	(0.043)	
second-party	-0.143***	0.87
	(0.046)	
unannounced (prior audit)	0.038	1.04
	(0.055)	
unannounced	-0.233***	0.79
	(0.053)	
number of recorded violations (prior audit)	-0.013***	0.99
	(0.003)	
new factory audit (prior audit) [1]	-0.014	0.99
	(0.065)	
follow-up audit (prior audit) [1]	0.053	1.05
	(0.044)	
follow-up audit[1]	0.052	1.05
	(0.057)	
female lead auditor (prior audit) [1]	0.024	1.02
	(0.038)	
female lead auditor[1]	-0.026	0.97
	(0.042)	
first time lead auditor (prior audit)	0.109**	1.12
	(0.052)	
first time lead auditor	0.076	1.08
	(0.049)	
total workers log (prior audit) [1,2]	0.048***	1.05
	(0.015)	
percent female workers (prior audit) [1]	0.137	1.15
	(0.128)	
factory tenure (prior audit) [1]	-0.007	0.99
	(0.006)	
Major violation	-0.177***	0.84
	(0.031)	
audit sequence	0.051***	1.05
	(0.016)	
FIXED EFFECTS		
violation vintage	Included	

Table E-3: Robustness results, alternative H2 measure (continued)

violation functional sub-category	Included
country	Included
industry	Included
audit year	Included
RANDOM EFFECTS	
audit	Included
factory	Included

Notes: *p<0.10, **p<0.05, ***p<0.01; N = 72,983 observations from 8,616 audits of 2,626 factories. Model is estimated using a multilevel mixed-effects logistic regression model with standard errors clustered at the factory. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model. The odds ratios (OR) presented in column 2 are calculated by exponentiating the coefficients in column 1.

Table E-4: Robustness results, predicting focal audit compliance

DEPENENT VARIABLE: # Violations	(1)	(2)
percent supplied (prior audit)[1]	-0.110***	-0.108***
	(0.025)	(0.039)
number of alternative suppliers (prior audit)	-0.009**	-0.009**
	(0.004)	(0.004)
second-party (prior audit)	0.068***	0.070***
	(0.011)	(0.016)
second-party	0.171***	0.169***
	(0.011)	(0.021)
unannounced (prior audit)	0.022	0.014
	(0.013)	(0.024)
unannounced	0.093***	0.099***
	(0.012)	(0.019)
new factory audit (prior audit) [1]	-0.117***	-0.091***
	(0.016)	(0.022)
follow-up audit (prior audit) [1]	-0.072***	-0.060***
	(0.011)	(0.017)
follow-up audit[1]	-0.040***	-0.043*
	(0.014)	(0.024)
female lead auditor (prior audit) [1]	0.012	0.006
	(0.010)	(0.016)
female lead auditor[1]	0.062***	0.060***
	(0.011)	(0.017)
first time lead auditor (prior audit)	0.040***	0.035*
	(0.012)	(0.020)
first time lead auditor	0.126***	0.126***
	(0.012)	(0.022)
total workers log (prior audit) [1]	-0.013***	-0.009
	(0.005)	(0.007)
percent female workers (prior audit) [1]	-0.028	-0.049
	(0.051)	(0.061)
factory tenure (prior audit) [1]	-0.006	-0.009*
	(0.007)	(0.005)
audit sequence	-0.038***	-0.010
	(0.007)	(0.010)
N (audits)	7,843	8,616
Number of factories	1,855	2,626
FIXED EFFECTS		
audit year	Included	Included
country		Included
industry		Included

Table E-4: Robustness results, predicting focal audit compliance (continued)

factory	Included	
RANDOM EFFECTS		
factory		Included

Notes: *p<0.10, **p<0.05, ***p<0.01; Models are estimated using a Poisson Regression. Model 1 includes factory fixed effects, and Model 2 includes country and industry fixed effects with factory random effects. Unit of analysis is at the factory audit. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model.

Appendix F – Analyses Supporting Identification Strategy

Table F-1: Frequency distribution of number of recorded violations (prior audit), by high and low supplier switching costs (H1)

Number of Recorded Violations (prior audit)	low supplier switching costs	high supplier switching costs	*Total*
1	142	123	*265*
2	244	231	*475*
3	284	313	*597*
4	327	322	*649*
5	353	360	*713*
6	343	327	*670*
7	315	317	*632*
8	269	304	*573*
9	267	289	*556*
10	187	226	*413*
11	201	197	*398*
12	171	173	*344*
13	134	149	*283*
14	103	117	*220*
15	102	105	*207*
16	58	82	*140*
17	47	79	*126*
18-76	354	376	*730*
Total	*3,901*	*4,090*	*7,991*

Notes: Chi2 test of independence yields a statistic of 78.84 with an associated p-value of 0.019. N = 7,991 unique factory audits in sample. 17 prior audit violations reflects the 90th percentile of the distribution.

Table F-2: Average violation severity, by high and low supplier switching costs (H1)

Switching Costs	Number of recorded violations (prior audit)	Major Violation
low supplier switching costs	13.8	83%
high supplier switching cost	14.0	84%
Two-tailed t-test p-value[1]	0.018	0.017
N	7,991	48,279
Unit of analysis	Factory-audit	Factory-violation

Notes: [1] A two-sample proportional test is used when testing for average differences in Major violations.

Table F-3: Distribution of violation observations across major functional categories, by high and low supplier switching costs (H1)

Violation Functional Category	low supplier switching cost	high supplier switching cost	*Total*
1. Employment Freely Chosen	100	81	*181*
2. Freedom of Association	828	794	*1,622*
3. Working Conditions	11,259	10,950	*22,209*
4. Environmental Regulations	492	477	*969*
5. Child Labor	620	612	*1,232*
6 Living Wages	3,831	3,799	*7,630*
7. Working Hours	4,431	4,278	*8,709*
8. Worker discrimination	74	70	*144*
9. Regular Employment	1,281	1,105	*2,386*
10. Punitive Worker Treatment	43	46	*89*
11. Legal Requirements	1,280	1,238	*2,518*
12. Worker knowledge of Code	287	303	*590*
Total	*24,526*	*23,753*	*48,279*

Notes: Chi2 test of independence yields a statistic of 12.068 with an associated p-value of 0.36. N = 48,279 unique factory violations.

Table F-4: Frequency distribution of number of recorded violations (prior audit), by high and low buyer switching costs (H2)

Number of Recorded Violations (prior audit)	Low buyer switching costs	High buyer switching costs	*Total*
1	137	128	*265*
2	231	244	*475*
3	316	281	*597*
4	359	290	*649*
5	389	324	*713*
6	399	271	*670*
7	363	269	*632*
8	323	250	*573*
9	315	241	*556*
10	229	184	*413*
11	235	163	*398*
12	180	164	*344*
13	150	133	*283*
14	112	108	*220*
15	116	91	*207*
16	71	69	*140*
17	68	58	*126*
18-76	364	366	*730*
Total	*4,357*	*3,634*	*7,991*

Notes: Chi2 test of independence yields a statistic of 80.24 with an associated p-value of 0.015. N = 7,991 unique factory audits in sample. 17 prior audit violations reflects the 90th percentile of the distribution.

Table F-5: Average violation severity, by high and low buyer switching costs (H2)

Switching Costs	Number of recorded violations (prior audit)	Major Violation
low buyer switching costs	13.2	84.1%
high buyer switching cost	14.5	83.4%
Two-tailed t-test p-value[1]	0.000	0.052
N	7,991	48,279
Unit of analysis	Factory-audit	Factory-violation

Notes: [1] A two-sample proportional test is used when testing for average differences in Major violations.

Table F-6: Distribution of violations across major functional categories, by high and low buyer switching cost (H2)

Violation Functional Category	Low buyer switching cost	High buyer switching cost	*Total*
1. Employment Freely Chosen	80	101	*181*
2. Freedom of Association	779	843	*1,622*
3. Working Conditions	11,141	11,068	*22,209*
4. Environmental Regulations	478	491	*969*
5. Child Labor	622	610	*1,232*
6 Living Wages	3,819	3,811	*7,630*
7. Working Hours	4,359	4,350	*8,709*
8. Discrimination	77	67	*144*
9. Regular Employment	1,162	1,224	*2,386*
10. Punitive Worker Treatment	42	47	*89*
11. Legal Requirements	1,297	1,221	*2,518*
12. Worker knowledge of Code	278	312	*590*
Total	*24,134*	*24,145*	*48,279*

Notes: Chi2 test of independence yields a statistic of 12.348 with an associated p-value of 0.34. N = 48,279 unique factory violations.

Table F-7: Distribution of prior audit violations, by high and low supplier switching costs (H1)

Number of recorded violations (prior audit)	Low supplier switching cost	High supplier switching cost	*Total*
1	185	195	*380*
2	274	319	*593*
3	306	375	*681*
4	344	360	*704*
5	356	425	*781*
6	372	333	*705*
7	325	345	*670*
8	289	311	*600*
9	282	297	*579*
10	216	209	*425*
11	200	203	*403*
12	173	183	*356*
13	152	140	*292*
14	103	119	*222*
15	116	95	*211*
16	74	68	*142*
17	60	68	*128*
18-76	382	362	*744*
Total	*4,209*	*4,407*	*8,616*

Notes: Chi2 test of independence yields a statistic of 64.35 with an associated p-value of 0.18. N = 8,616 unique factory audits. The unit of analysis is the factory audit because prior audit violations and supplier switching costs are determined at the factory audit. 17 prior audit violations reflects the 90th percentile of this distribution.

Table F-8: Regressing focal audit supplier switching costs on prior audit violations (H1)

DEPENENT VARIABLE: percent supplied (focal audit)	(1)
Number of recorded violations (prior audit)	-0.001
	(0.000)
number of alternative suppliers (prior audit)	0.002
	(0.002)
second-party (prior audit)	-0.006
	(0.005)
second-party	-0.004
	(0.006)
unannounced (prior audit)	0.010
	(0.007)
unannounced	-0.001
	(0.007)
new factory audit (prior audit) [1]	-0.025***
	(0.007)
follow-up audit (prior audit) [1]	0.003
	(0.005)
follow-up audit[1]	0.005
	(0.008)
female lead auditor (prior audit) [1]	0.005
	(0.005)
female lead auditor[1]	-0.001
	(0.005)
first time lead auditor (prior audit)	-0.007
	(0.006)
first time lead auditor	0.005
	(0.006)
total workers log (prior audit) [1]	0.009***
	(0.003)
percent female workers (prior audit) [1]	-0.026
	(0.034)
factory tenure (prior audit) [1]	-0.008*
	(0.005)
audit sequence	0.010*
	(0.005)
N (audits)	8,616
Number of factories	2,626
FIXED EFFECTS	
audit year	Included
factory	Included

Table F-8: Regressing focal audit supplier switching costs on prior audit violations (H1) (continued)

Notes: *p<0.10, **p<0.05, ***p<0.01; The model is estimated using a linear regression and includes factory fixed effects. Unit of analysis is the factory audit. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model.

Appendix G – Additional Tables and Charts for Chapter 3

Table G-1: Data pipeline

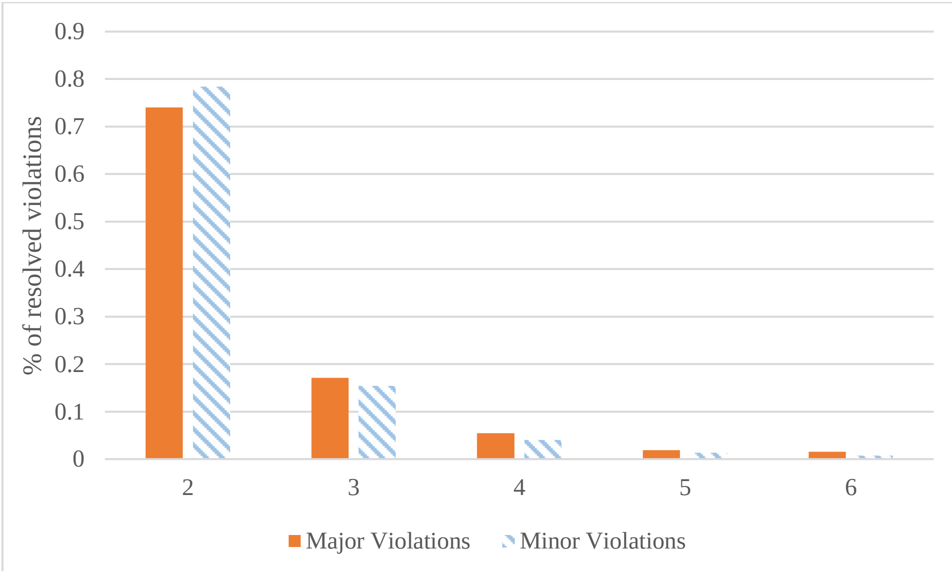
Description		Total number of factory-audit-violations	% loss in data from prior row
All factory-audit-violations	[1]	185,904	
Omit factory's first audit	[2]	116,444	37.4%
Omit factory-violation's first appearance in the data	[3]	72,983	37.3%

Notes:

[2]: I omit a factory's first audit as the model includes several prior audit predictors, which would be zero in a factory's first audit.

[3]: I omit the factory-violation first appearance in the data because violations cannot be resolved in the same audit in which they were discovered (the probability of resolution equals 0 for the violation's first audit).

Figure G-2: Distribution of Resolved Violations by the Number of Audits it Takes to Resolve



Appendix H – Detailed Description of Future Research

I describe below how text data has been used in prior strategy and operations research and then describe how it might be applied to the text in this dataset.

Analyzing textual data in management research is becoming a more popular trend in many domains, including strategy and operations management.[98] These domains use a variety of natural language processing (NLP) techniques, such as sentiment analysis. Both strategy scholars and scholars in operations management research tend to use NLP techniques to help measure various concepts (Kang et al., 2020). For example, Cui et al (2018) used text in Facebook comments of a men's clothing online retailer's page to test whether the level of endorsement in Facebook comments could improve daily sales forecasts. To measure comment endorsement, they measured two aspects of each comment, the comment's informativeness and sentiment. As prior scholars have done, they measured informativeness by "the number of words, sentences, and unique words in a comment" (Cui et al., 2018; 6). To measure a comment's sentiment, they used a recursive neural net method and pre-defined corpus to extract a vector of each phrase in the comment. Through this method, the comment was then classified into positive, negative, or neutral sentiment. Daily comments on each sentiment were totaled and the authors then used this metric, along with the informativeness metric, in a model which predicts the company's daily sales. They found that the inclusion of these metrics significantly improved daily sales forecasts.

Two potential directions for future research are described below that rely on the textual information contained in this dataset. Similar to how NLP has been used in prior operations

[98] See Kang et al. (2020) for a review of different applications and approaches of NLP in top management journals by research domain.

management and strategy papers, these directions rely on using the text to construct different measures that predict violation resolution. First, one easier application would be in line with Cui et al (2018) – that is, to construct a measure of corrective action informativeness to explore whether corrective actions that are measured to be more informative are associated with greater odds of violation resolution. This measure would be relatively simple to construct after cleaning the text data (e.g., stemming and removing HTML characters) by counting the number of unique words or phrases in the corrective action. One potential hypothesis could be that comments that are more informative (by containing more unique words or phrases) increase violation resolution odds because they contain better instruction on how to resolve a particular violation. This is related to prior literature that relates workers receiving more specific feedback to better future performance because more specific feedback can provide useful information for correcting incorrect approaches to tasks (Bandura, 1991; Bandura & Cervone, 1983).

A second direction, which is also in line with Cui et al (2018), would be to perform sentiment analysis. This could be useful in a hypothesis that predicts that corrective actions that have more positive sentiment are associated with greater odds of violation resolution. This would be because some studies have found that feedback can hamper performance when it's seen as negative (Fedor, Davis, Maslyn, and Mathieson, 2001), overly critical, or controlling (Baron, 1988, 1990, 1993). Alternatively, one could connect corrective actions with a positive sentiment as being reflective of the auditors having a more cooperative auditing approach, which has been previously linked to improved compliance but through different measures (Locke, Amengual, and Mangla, 2009).

Sentiment analysis can potentially be utilized to measure variables that tests the above hypothesis in the following way. First, one would represent each corrective action as a vector.[99] Second, an algorithm would score each corrective action vector by its sentiment using either a pre-defined dictionary or a user-defined dictionary, which will link specific words or phrases into a score for a given sentiment.[100] Pre-defined dictionaries are quicker to access and use, but their sentiment prediction accuracy tends to be better when assessing text from the subject domain that the dictionary was trained on (e.g., a sentiment dictionary that was trained using movie reviews will most accurately predict the sentiment of movie reviews).[101] Therefore, it's possible that using a pre-defined dictionary could result in low accuracy sentiment predictions unless the dictionary was trained on similar auditing corrective action data (and I am not aware of any such dictionary). Once a dictionary has been selected it would be used to assign a sentiment score to each corrective action vector. This score could, for example, measure the portion of corrective action words that are positive, and this measurement could be used to predict violation resolution propensity.

---

[99] Cui et al (2017) uses a "recursive neural tensor network (RNTN) on top of the Stanford Sentiment Treebank corpus to extract a compositional vector representation of each phrase in the comment" (pg 6) as is done in a seminal paper on natural language processing (Socher et al. 2012).

[100] Some popular pre-defined sentiment dictionaries are: Stanford sentence and grammatical dependency parse, Sentistrength, Linguistic inquiry and word count (LIWC), and Soutlab/lithium social media sentiment monitoring systems.

[101] An alternative is to create a dictionary from the corrective action words or phrases. This can be done by the researcher (but would be time intensive) or can be outsourced to a platform like MTurk. To create a dictionary, one could classify the data into a training and testing set, label the words in the testing set according to its sentiment, and then predict the sentiment on the testing set (to assess accuracy).

Appendix I – Alternative Industry Construction and Analysis

To manually construct supplier industries, I used three sources of data. First, I imputed a supplier's industry for 1,060 (40%) suppliers in the sample using the products listed in their audit report.[102] Second, I imputed 947 (36%) supplier industries using the supplier's name.[103] Third, I web-searched the remaining 619 (24%) of suppliers and identified their industry using information on the supplier's webpage. The three industries with the greatest number of suppliers are clothing garments (56%), footwear (9%), and apparel accessories (7%).

I next assessed whether the main model's results were robust to the re-measurement of *number of alternative suppliers (prior audit)* using the re-constructed supplier industry categories, as that would provide an indication of a positive correlation between the measure of *number of alternative suppliers (prior audit)* constructed using product codes to define industries (as is done in the main analysis) and the alternative measure constructed using the alternative supplier industries. Estimating a model using the alternative variable, *number of alternative suppliers (prior audit)'*, yields similar results to the main model that tests H2, and those results are located in Table I-1.[104]

Finally, I compare the share of suppliers by industry in 2012 located in China, the country with the greatest share of suppliers and audits, to the share of enterprises by industry as provided by the China Data Online.[105] Located in Table I-2, I find that the share of suppliers in

[102] I was not able to use the products listed in the audit report to impute industry for all suppliers in the data either because I do not have access to all supplier audit reports in the database or because the products field is blank in a supplier's audit report.
[103] For example, I assume a supplier with the word "cosmetics" in the name should be mapped to the "cosmetics" industry.
[104] The alternative measure was constructed in an identical fashion to the variable used in the main analysis except that it uses the alternative industry categories.
[105] Enterprise reflects a number of different company ownership categories: those that are state-owned, privately owned, cooerpatives, collectively owned, those that are shareholding enterprises, and those with funds from foreign entities.

TrendyStyle's data appears to be similar to the share of suppliers in the China Data Online data. One notable exception is that the share of suppliers in "Textiles and Clothing, Apparel Industry" is 9% higher in the TrendyStyle data than in the external data source, which is likely due to TrendyStyle itself primarily selling apparel, so the brand will naturally concentrate more of its sourcing to apparel-related suppliers.

Table I-1: Robustness results using alternative H2 measure

DEPENDENT VARIABLE= Resolved	(1)	OR
percent supplied (prior audit)[1]	0.222***	1.25
	(0.069)	
number of alternative suppliers (prior audit)'[2]	0.279***	1.32
	(0.043)	
second-party (prior audit)	-0.076*	0.93
	(0.043)	
second-party	-0.139***	0.87
	(0.045)	
unannounced (prior audit)	0.037	1.04
	(0.054)	
unannounced	-0.228***	0.80
	(0.053)	
number of recorded violations (prior audit)	-0.012***	0.99
	(0.003)	
new factory audit (prior audit) [1]	-0.082	0.92
	(0.061)	
follow-up audit (prior audit) [1]	0.065	1.07
	(0.044)	
follow-up audit[1]	0.052	1.05
	(0.057)	
female lead auditor (prior audit) [1]	0.020	1.02
	(0.038)	
female lead auditor[1]	-0.019	0.98
	(0.042)	
first time lead auditor (prior audit)	0.098*	1.10
	(0.051)	
first time lead auditor	0.071	1.07
	(0.049)	
total workers log (prior audit) [1,2]	0.054***	1.06
	(0.016)	
percent female workers (prior audit) [1]	0.097	1.10
	(0.136)	
factory tenure (prior audit) [1]	-0.007	0.99
	(0.006)	
Major violation	-0.174***	0.84
	(0.031)	
audit sequence	0.061***	1.06
	(0.015)	
FIXED EFFECTS		
violation vintage	Included	

Table I-1: Robustness results using alternative H2 measure (continued)

violation functional sub-category	Included
audit year	Included
RANDOM EFFECTS	
audit	Included
factory	Included

Notes: *p<0.10, **p<0.05, ***p<0.01; N = 72,983 observations from 8,616 audits of 2,626 factories. Model is estimated using a multilevel mixed-effects logistic regression model with standard errors clustered at the factory. [1] denotes variables where missing values are recoded to zero and model includes dummy variables to indicate instances in which variables were missing data and thus recoded to 0. [2] denotes variable where natural log is used in model. The odds ratios (OR) presented in column 2 are calculated by exponentiating the coefficients in column 1.

Table I-2: Comparison of 2012 company share in China

Industry	**Source: China Data Online, Yearly Statistics (2012)** # of enterprises	**% share**	**TrendyStyle Database** # of suppliers	**% share**
Textiles and Clothing, Apparel Industry	14,788	50%	233	59%
Cosmetics	336	1%	6	2%
Glasses	216	1%	7	2%
Plastic Furniture	74	0%	1	0%
Leather Luggage and Handbags	1,209	4%	10	3%
Household Textiles	1,786	6%	21	5%
Jewelry and Related Articles	403	1%	8	2%
Ceramics	1,848	6%	8	2%
Paper Products	4,201	14%	36	9%
Shoes with Textile Outside, Leather, or Plastic	3,446	12%	55	14%
Toy	1,266	4%	11	3%
TOTAL	*29,573*		*396*	

Notes: Relevant manufacturing industries manually identified from China Data Online. Industry definitions taken from China Data Online and industries in TrendyStyle database are mapped to China Data Online industry definitions. Data from TrendyStyle Database reflects the number of suppliers in the data sample that are located in China and audited by TrendyStyle in 2012. In the China Data Online, enterprise reflects a number of different company ownership categories: those that are state-owned, privately owned, cooerpatives, collectively owned, those that are shareholding enterprises, and those with funds from foreign entities.