Responsible Sourcing: The First Step Is the Hardest

Pia Ramchandani, Hamsa Bastani, Ken Moon
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104
{piar2, hamsab, kenmoon}@wharton.upenn.edu

By sourcing responsibly, companies promote corporate social responsibility (CSR) throughout their global supply chains. However, most consumable good firms' product lines contain only a few targeted products certified by Fair Trade or other third-party certifiers, which suggests that CSR initiatives are limited. We instead present evidence that certifications enable firms to learn *how* to source responsibly. By successfully certifying even a single product, companies may enjoy positive, knowledge-based spillovers encouraging responsible sourcing throughout their product lines. Using data tracking the sourcing decisions of US consumer coffee companies ($48B market) and their CSR violations reported in global disclosures and news sources, we find that firms' CSR violation rates drop upon certifying—but surprisingly, by a similar magnitude whether the resulting penetration of certified products within a firm's product line is high or low, consistent with initial certifications enabling disproportionately large CSR improvements through firm-level learning. Consequently, incremental certifications by firms with little to no previous certifications realize 46-fold larger reductions in violations over equivalent expansions made by firms with substantial certified sales. In contrast, companies without certifications show no improvement in their CSR violation rates upon making their own reputation-backed CSR claims. We further examine the degree to which country-based sourcing mediates certification's effects. Dual-sourcing companies achieve significantly lower CSR violation rates even when sourcing from countries in which violations are frequent. Our work suggests that prevalent dual-sourcing amplifies, rather than limits, responsible sourcing in supply chains, and that certified sourcing uniquely develops the pool of responsible suppliers in high-risk countries.

Key words: Certifications, Corporate social responsibility, Human rights, Labor abuse, Sustainable sourcing

1. Introduction

Over recent decades, consumers, investors, governments, and stakeholders have intensified scrutiny on the social and environmental impacts of firms' business operations. As firms increasingly source across borders, significant concerns about the abuses taking place within their global supply chains have followed suit. According to the International Labor Organization, global supply chains' 450 million workers include 168 million subject to child labor, 21 million victims of forced labor, and 12 million annual deaths from unsafe work environments (ILO 2018). In environmental impact, supply chains for consumer goods contribute to over 80% of water loss, 60% of greenhouse gas emissions, and 67% of deforestation (Consortium 2016).

In response to such abuses, companies have created corporate social responsibility (CSR) initiatives. Such initiatives voluntarily undertake to identify and curb the harmful effects of supply chain activities (Lee and Tang 2018, Li and Wu 2017). Firms source responsibly by committing to working with suppliers to enforce CSR standards, which can include minimizing child labor, improving working conditions, and decreasing environmental impacts such as deforestation.

However, actually sourcing responsibly is operationally challenging, requiring that firms enforce well-functioning governance, visibility, and transparency throughout their global supply chains (Locke et al. 2007a, Pagell and Wu 2009, Sodhi and Tang 2018). Firms must develop ethical standards for supplier behavior, disseminate them to diverse suppliers, and then audit supplier practices, either internally or by engaging third-party certification agencies such as Fair Trade (Thorlakson et al. 2018). Finding or developing reliably responsible suppliers requires overcoming considerable geographical distances and cultural and organizational divides, making local knowledge valuable (Boström 2015). The lack of generally agreed-upon ethical standards further impedes the coordination of sourcing and audits (Nelson et al. 2018). Even then, suppliers may illegally subcontract (Caro et al. 2018) or deliberately evade ethical responsibilities by sourcing from regions with looser regulations (Abbasi 2017, Hasle and Jensen 2012). Finally, to generate market demand, firms must persuasively communicate their sourcing to socially responsible consumers through marketing and on-package labeling (Kraft et al. 2019, Buell et al. 2019).

On the other hand, once firms develop the requisite operational and institutional knowledge to enforce and monitor ethical behavior, they may enjoy diminished marginal costs when further expanding the scope of their responsible sourcing to additional products. This paper examines whether companies that certify just a few products as part of their product lines accrue learning-based spillovers that enable them to also reduce violations in the non-certified sections of their supply chains. This question is especially relevant because dual-sourcing—which for our purposes, we define as a firm sourcing from both certified and non-certified supply channels—is a common sourcing strategy (Thorlakson et al. 2018). Nielsen RetailScanner data show that whereas 21% of grocery brands dual-source, only 5% of grocery sales are actually certified. Thus, if only third-party certified goods are reliably responsible, the combination of prevalent dual-sourcing with actually low certification levels should cause concern.

We address this empirical question using data matching the certification decisions of firms in the $66B US consumer coffee and chocolate markets with the violations of labor and environmental standards reported in their supply chains. Our key empirical finding shows that, as firms initially certify a first few products, they appear to accrue spillover-inducing gains in operational knowledge about *how* to source responsibly. Specifically, when coffee and chocolate companies certify even small parts of their product portfolio, namely above 3% in sales, they accrue positive spillovers

enabling ethical sourcing throughout the uncertified rest of their portfolios. While firms' CSR violations significantly drop upon certifying any products, once they certify, their violation rates remain low and unexpectedly steady regardless of whether their product lines are, illustratively, 5% certified or 100% certified. By showing that companies accrue nearly all of their CSR improvements upon crossing very low certification thresholds, our results convey that certifying most or all of their products is not a necessary mechanism for firms to source responsibly. Instead, CSR gains arrive disproportionately upon firms' initial certifications, which is particularly consistent with positive spillovers based on the firms' institutional learning from certifying a first product. Importantly, such learning can occur both through firms' direct interactions with experienced third-party certifiers or independently through their own institutional efforts aimed at meeting certification standards. In magnitude, we estimate that marginally certifying products among firms with product-line penetration below 3% certified can reduce violation outcomes by 46× more than for heavily certified companies. Such empirical findings draw on panel data which tracks each firm's marginal response in violations to changes in its level of product-line certifications over time.

We address a number of alternative explanations. Hypothetically, companies could institute CSR improvements widely at scale, thus lowering violations, followed gradually by the formal certifications of qualified products. While the act of batching CSR efforts is notably consistent with the economics of learning effects, firms also empirically fail to exhibit the monotonically increasing certification patterns consistent with lagged certifications. As a second alternative explanation, ethical firms may perform similarly well but differ in their signalling strategy (e.g, some primarily certify while others conduct their own programs). However, this is unlikely because it suggests that the sole role of certification is signaling, which does not align with existing research or our findings on companies with reputation-backed CSR claims. Under a third alternative mechanism, firms may rely on certifications to replace their worst-performing suppliers first, such that their certifications sharply reduce CSR violations initially but yield decreased marginal returns thereafter. However, our findings would then require that all dual-sourcing companies initially source exactly 3% of their production from extreme offenders, which is unlikely given the diversity of company sizes, mixes of international sourcing, and pre-certification rates of CSR violations represented by the firms in our data sample.

To further examine certification's effects and corroborate our learning hypothesis, we expand the analysis to account for the studied firms' country-based sourcing. We collect the countries which each coffee or chocolate companies sources from using import shipments data gleaned from bills of ladings. By controlling for firms' countries of sourcing, we remove it as a mediating factor that could explain, in part or in whole, certification's effect on CSR performance. We similarly study

the mediating role of geographical sourcing on the CSR performance of firms that do not certify, but make independent on-packaging CSR claims backed solely by their own firm reputations.

We find that companies that make reputation-based claims do *not* exhibit lower violations rates. In contrast, companies with more than 3% in certified sales maintain low violations even when sourcing from high-risk countries. This suggests either that they work with certification agencies to rigorously source ethical products from these regions, or that they develop the necessary internal capabilities by working with the agencies elsewhere. Thus, whereas both groups of firms are motivated to source responsibly, learning to source rigorously appears to uniquely originate with certifications. Putting aside what firms learn, our findings show that certifications play a vital role in developing the supplier base in high-risk regions. In this way, certifications uniquely impact the extensive margin of socially responsible production.

Our work makes several contributions. First, whereas an extensive literature raises concerns regarding the prevalence of dual-sourcing, we examine an under-appreciated benefit from partially certifying a firm's product line. By enabling firms to learn the operational practices behind sourcing responsibly, even limited certification efforts generate spillover effects that allow firms to reduce CSR violations in their uncertified sourcing channels. Thus, we novelly argue that dual-sourcing amplifies, rather than limits, responsible sourcing in supply chains. Second, the literature remains divided on whether firm reputations adequately ensure the credibility of firms' product labeling, especially when firms place CSR claims on their product packaging without certification (Giovannucci et al. 2008, Darnall and Sides 2008). Our results indicate that reputation-based claims do not appear to deliver, even when CSR-friendly sourcing options are readily available to firms. In addition, our evidence supports that certified companies undertake the additional challenge of sourcing responsibly from high-risk countries. Lastly, our findings inform the role of CSR certification programs. To maximize impact on CSR outcomes, certification agencies should leverage spillovers by emphasizing the on-boarding of new firms through first-time certifications and expand the responsible supplier pool in high-risk countries.

1.1. Related Literature

Our work contributes to the literature regarding the role of third-party certifications in CSR. The efficacy of a certification depends on customers seeking its affiliated products. Therefore, in product markets where consumers stand to directly benefit from and commonly care about good behavior, such as with concerns about food safety or the cost savings received from energy-efficient household appliances, certifications have proven an effective mechanism through which to enforce CSR (Waldman and Kerr 2014). However, when it comes to ethical fashion apparel and sustainably caught seafood, for example, certified products appeal mainly to a select group of

socially responsible consumers. In such cases, theory suggests that dual-sourcing can detriment CSR outcomes by segmenting the consumer market (Guo et al. 2015). Without necessarily improving CSR outcomes at all, dual-sourcing may superficially appease consumers, who often exhibit similar willingness to pay for small or heavy investments in CSR (Trudel and Cotte 2009). Empirically, dual-sourcing is troublingly prevalent (Thorlakson et al. 2018).

An extensive literature examines the critical role of supply chain transparency and audits in achieving supplier compliance with CSR. Supply chain transparency entails that a firm disclose information to consumers, investors, and other stakeholders about its CSR activities and compliance. As such, it goes beyond, but necessitates, visibility, which is the firm's own monitoring of activities and compliance in its supply chain (Sodhi and Tang 2018, Kraft et al. 2018). Joint audits and supplier list disclosures can beneficially penalize suppliers for their non-compliant behaviors (Kalkanci and Plambeck 2018b, Chen et al. 2019, Fang and Cho 2019) or induce appropriate monitoring by outside organizations such as NGOs (Chen et al. 2018). Kim (2015) studies how regulators can increase transparency by encouraging firms to voluntarily disclose violations.

At the same time, several pitfalls can harm transparency-based efforts. Disclosure mandates can deter firms from learning about supplier violations (Kalkanci and Plambeck 2018a). Overly frequent audits may cause suppliers to direct their efforts towards evading the audits rather than actually preventing harm (Plambeck and Taylor 2015), and suppliers under pressure may engage in unauthorized subcontracting (Caro et al. 2018). CSR progress can also be impeded when it is difficult for firms to credibly signal responsibility, *i.e.*, transparency is costly (Guo et al. 2017).

Yet, a growing body of research suggests that CSR compliance is enhanced by collaborative relationships with suppliers which constructively address managerial challenges and standards. Porteous et al. (2015) find that incentivizing supplier compliance through increased business, public recognition, and training strongly associates with reduced supplier violations. Corbett et al. (2005) and Castka et al. (2015) demonstrate that adopting effective quality management and environmental standards leads to improved financial performance. In a project with Nike, Locke et al. (2007b) illustrate that whereas monitoring suppliers generally fails to improve poor factory working conditions, the conditions do improve when suppliers are given assistance in identifying and addressing operational root causes (*e.g.*, problematic scheduling practices). Amengual and Distelhorst (2019) and Boudreau (2020) study the role of management practices in improving factory labor and safety compliance, and Bloom et al. (2010) demonstrate that better managed firms emit significantly less greenhouse gas. Distelhorst et al. (2015) and Toffel et al. (2015) emphasize the role of the surrounding social and regulatory institutions in compelling compliance, while Bastani and de Zegher (2019) find that banning supplier practices related to harmful behavior may reduce violations. Akkaya et al. (2020) study learning-by-doing as a hurdle for suppliers in adopting sustainable practices.

Finally, contractual arrangements and market institutions can facilitate CSR. By incentivizing community monitoring and enforcement, de Zegher et al. (2018) propose contracts to lessen deforestation in palm oil supply chains. Levi et al. (2019) connect agricultural wholesale marketplaces onto a common platform to improve small farmer welfare. Babich and Tang (2012) study inspection and market response as two alternative forms of contractual discipline, and Chen and Lee (2016) and Cho et al. (2019) design optimal contracts that combine supplier screening mechanisms and contingency payments. Additional research on contracts incorporates dynamics and multiple tiers (Huang et al. 2017, Lewis et al. 2019, Zhang et al. 2019). Orsdemir et al. (2019) focuses on supply chain structure and integration.

The remainder of the paper proceeds as follows. Section 2 describes how we sourced and matched our dataset, as well as its limitations. Section 3 discusses sourcing spillovers, suggested both by data and first-hand sources. Section 4 covers our empirical methodology and results, and Section 5 expands on implications.

2. Industry and Data

Our primary data cover firm sourcing decisions and metrics of firms' supply chain responsibility in the $48B US coffee market from 2006-2016. We additionally corroborate our main findings using analogous data from the chocolate market. The data integrate nearly 1TB of data from several sources to report company sourcing strategies and violations over time.

Choice of industry. We study the coffee industry for three reasons. First, coffee companies are highly developed and differentiated in terms of sourcing behavior. Coffee has a well-established history of certifications, as shown in Figure 1 for the number of coffee products certified by Fair Trade. Additionally, coffee includes significant representation for companies making their own reputation-backed claims, not backed by third-party certifications, around ethical sourcing (see Figure 2(b)). We study whether such firms' products are responsibly sourced in comparison to certified products. Academically, it is important to understand the extent to which reputation alone can motivate companies to police supplier CSR, and the comparison also sheds light on the advantages, if any, of certifications. Relevantly, independent reputation-backed CSR claims continue to gain prevalence, with 31% of global grocery sales in 2016 consisting of products featuring such claims on their packaging.

Second, coffee production has been marred by significant ethical lapses, making the problem of responsible sourcing highly relevant in practice. At present, many coffee plantations fail to abide by labor laws, provide poor and hazardous working conditions, and exploit migrant and child laborers (Zamora 2013, US Department of Labor 2013). Regarding coffee's environmental impact, millions of acres of forest have been cleared illegally, destroying natural habitats (WWF 2007, The Guardian 2011).

Finally, coffee products delineate into relatively homogenous specialty and commodity product segments, which allows us to more effectively control for product attributes. Certified goods virtually always fall into the specialty segment. Because coffee certifications address only the subset of consumers who are socially responsible, dual-sourcing is empirically prevalent in the coffee market, which is of interest for our study.

To corroborate our findings, we replicate our analyses in the US consumer market for chocolate. While the chocolate industry has experienced similarly high ethical violation rates, it lacks coffee's long history with certification programs, with cocoa certifications becoming popular only in the past decade (see Figure 1).

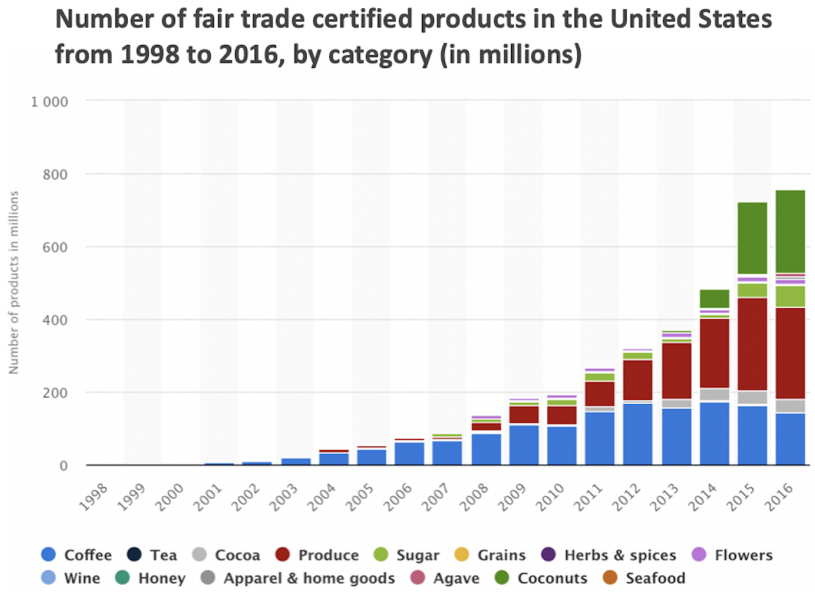


Figure 1 **Fair Trade certified products by category and year**

(a) Dual-sourcing product portfolio with a portion of certified products

(b) Products with self-made claims (with no 3rd party certification)

Figure 2 **Coffee Product Packaging with Sourcing-related Labels**

Data sources. To measure firms' supply chain CSR performance, we use the incidence of CSR violations in their supply chains. Data regarding the studied firms' environmental, social and corporate governance (ESG) violations are sourced from RepRisk, which is a data provider that, on a daily basis, aggregates and cross-checks negative ESG incidents identified by over 80,000 news, media, regulatory, and commercial information sources published in over 20 languages. The violations data coverage is comprehensive, recording and classifying all detected violation events for more than 140,000 companies over the span of 2006 to 2016.

Regarding the firms' sourcing strategies, we source data from Nielsen Label Insights, which is derived from consumer product packaging and labels from 2006-2016. The data provide label attributes granularly classifying each UPC-level product's on-packaging claims and certification endorsements. From the coffee products' individual claims and certifications, we aggregate by company to construct firms' product line portfolios (*e.g.*, a product line may be 30% certified, 15% independent claims, and 55% with no CSR-related labeling). Hence we reconstruct firms' sourcing decisions at the level of their product lines, which we can then link to their CSR performance.

As our last link of primary interest, we obtain data on the firms' first-tier, cross-border suppliers from the Panjiva platform. Panjiva extracts data from import documents (*e.g.*, bills of lading) covering over one billion transactions by over 8 million companies across 195 countries. Thus we identify the locations of the final international stop in the firms' coffee supply chains before the sourced coffee enters the US.

At times, it will be appropriate for us to normalize firms' detected counts for CSR violations. Two pieces of data are useful for this purpose. First, the Nielsen Retail Scanner dataset supplies product-level retail sales data, using which we derive firms' coffee-based revenues and unit sales. Thus we control for the fact that a firm's CSR violations may scale with the volume of production activity in its supply chain. Second, we use data from LexisNexis, one of the world's largest electronic databases for legal and journalistic documents, to measure the volume of general news related to each company during each year from 2006 to 2016. For the subset of our analyses that rely on cross-sectional matching across companies, we use these data to adjust firms' observed violation counts to account for their volumes of product sold, their speciality-versus-commodity mixes as captured by price points, and the levels of news scrutiny they receive. These adjustments are unnecessary for our main analyses, which include company fixed effects and hence produce estimates based on firm-specific variation in sourcing over time.

The final dataset contains 398 observations for 47 firms where each observation represents one year of data for one particular coffee company during the time period 2006-2016. Some companies were launched during this timeframe, and therefore only contain data for a subset of the years. Each observation contains information on the number of ESG violations, country-based sourcing,

scale of production, and news visibility for a particular company in that year. The final merged dataset covers 93% of certified product sales in the US consumer coffee market and 59% of all coffee sales (please see Appendix C for more information on the data merge and final sample size). Thus, we view the dataset as containing nearly all US companies with CSR-related labeling and a healthy selection of the companies most comparably positioned in marketing and sales (*e.g.*, specialty firms) among those making no CSR-related claims.

2.1. Outcome Variable

2.1.1. ESG violation events. Like Li and Wu (2017), our primary outcome variable is the (severity-adjusted) supply chain ESG violation counts for the company-year collected by RepRisk. At the time of our data extraction, RepRisk was screening over 500,000 documents every day across 80,000 public sources and stakeholders from over 130 countries in 20 languages to identify violations. Public sources include print media, online media, social media, government bodies, regulators, think tanks, newsletters, and other online sources. These sources include coverage at the local, national, regional, and international levels. RepRisk uses a combination of machine learning and manual review to tag incidents in 28 violation categories.[1]

Since we focus on responsible sourcing, we filtered events by the available ESG violation categories to only include those related to sourcing, *e.g.*, we do not consider events solely relating to tax evasion or executive compensation. Table 1 lists the violation categories used in our study.

In cross-sectional analyses, we control for differences in violation detection rates based on company news visibility and size. Because different companies source from different locales, we additionally recognize that the definitions of violations can effectively differ across jurisdictions. Thus, certain activities may not be defined as violations simply because they occur in a country with looser regulations and lower standards. First, this classification issue is addressed when companies are uniformly held to international or norm-based standards for ESG issues. Such standards are increasingly common and supported by the monitoring of NGOs and the media. Second, it is expected that irresponsible supplier behavior occurs more prevalently in loosely regulated nations. Because loose regulations relatively undercount violations, we would be concerned that the differences between countries may be washed out. However, when we rank our data sample's countries by violation rates, the order reassuringly accords with the views of expert observers (e.g., we find high violation rates from Indonesia and Papa New Guinea, and low violation rates from Peru).

One may be concerned that dual-sourcing companies source primarily from countries that underreport violations, and thus, have seemingly low violation rates regardless of certification coverage

[1] RepRisk's methodology is event-driven rather than targeting a specific set of companies; therefore, there is no censoring of companies that were not in operation for some subset of the years.

due to lack of detection. Therefore, we conduct simulations (see Appendix D) of varying levels of detection bias across countries to understand the impact on observed company violations. We find that detection bias in the violations data cannot plausibly explain the L-shape pattern we observe without learning-based spillovers.

Lastly, we carry out corroborating regressions that control for companies' country-based sourcing.

We therefore analyze 2,216 eligible ESG events occurring between 2006-2016 for the companies in our study. Furthermore, events are scaled by severity based on RepRisk's review on the impact of the event to avoid equal weighting for minor and major issues. For each company-year data observation, we aggregate the firm's detected violations (scaled by severity) into an event count reflecting its supply chain CSR performance.

Category	**Issue categories**	**% of Events Tagged**
Environment	climate change, GHG emissions, global pollution, impacts on landscapes, ecosystems, and/or biodiversity, local pollution, overuse and wasting of resources, waste issues	62%
Social	child labor, forced labor, human rights abuses and corporate complicity, impacts on communities, occupational health and safety issues, poor employment conditions	77%
Governance	violation of international standards, discrimination in employment, freedom of association and collective bargaining, social discrimination	17%

Table 1 ESG issue categories that qualify firms' violation events for inclusion in the study

Each studied violation event falls into one or more ESG issue categories listed above. An event can be tagged as a violation in multiple categories, so the % column adds up to more than 100%.

2.2. Study Variables

2.2.1. Sourcing strategies for CSR. We examine firms' sourcing strategies at the product line level over time. A firm's sourcing strategy is jointly characterized by the type and the extent of its ethical sourcing. We first classify firms into three types of CSR-related sourcing strategies, which are defined based on the sourcing composition of a firm's product line:

1. **Certified or dual-sourcing firms** certify a non-trivial fraction of their product lines through a major third-party certifier for ESG-related sourcing. For reasons explained in Section 3, we set the threshold for non-trivial certification at 3% of sales. These firms may also have additional products with self-made claims.

2. **Firms making reputation-backed claims** do not fall into category 1 but still label one or more products with claims regarding ESG-related sourcing. To accomplish responsible sourcing without third-party auditing or accreditation, these companies must develop their own internal

sourcing standards and programs. In practice, they may outsource audits or other aspects of program implementation, but the programs may enforce unaccredited standards. Some firms' claims may be duplicitous or misleading.

3. **Firms with no responsible sourcing claims** place no ESG-related labels on their products.

Within each sourcing type, we further consider the *extent* to which firms source responsibly. For example, a major question in our study is whether CSR performance improves when dual-sourcing firms increase the fraction of certified products in their product lines.

To determine a firm's sourcing type, we classify the products in its product line. We classify the data's 4,534 unique coffee products into "certified", "reputation-backed claims", and "no ESG labeling" as follows. From the Nielsen LabelInsights data reporting claims and endorsements placed on product packaging at the UPC level, we filter down to labels relating to ethical sourcing (*e.g.*, Fair Trade or Rainforest Alliance). To focus on CSR and ESG compliance, we omit certifications such as the Organic standard, which solely concern growing practices. We identify 12 unique formal certification program labels (see list in Appendix A) and 176 reputation-backed labels conveying ESG claims (e.g., as in Figure 2b). If a product bears a major certifier's third-party endorsement label, we classify it as a certified product. Since companies are incentivized to market towards socially responsible consumers, we assume that third-party certified products are visibly labeled. If a company labels a product with an uncertified ESG sourcing claim, the product is classified as making reputation-backed claims. All other products are designated as bearing no ESG labeling.

Of the 4,534 products, 13% are certified and 14% make reputation-backed claims. By cross-referencing with the Nielsen Retail Scanner pricing data, we confirm that the products certified as responsibly sourced are mainly categorized as specialty coffee. Consistent with the literature, certified products are priced at $9.8 on average, products with reputation-backed claims are priced at $9 on average, and the remaining products are priced at $8 on average. Figure 3 shows histograms of the categories' respective product counts and retail prices (per ounce).

2.2.2. Country-based sourcing. The Panjiva platform provides supplier shipment data for the set of companies matched between Nielsen Label Insights, Retail Scanner, and RepRisk for the study period. Each company file reports every first-tier overseas supplier from which the company received an international shipment, the number and contents of the shipments, and the supplier locations. We manually remove all non-coffee shipments, *e.g.*, of ancillary products such as straws or cups; we perform analogous filtering for chocolate shipments in our supporting analyses. We further distinguish between shipments received from coffee-growing nations and those received from middlemen (primarily North American and European) in the coffee supply chain. Figure 4 maps the sources of US coffee import shipments received from the coffee-growing countries in our data.

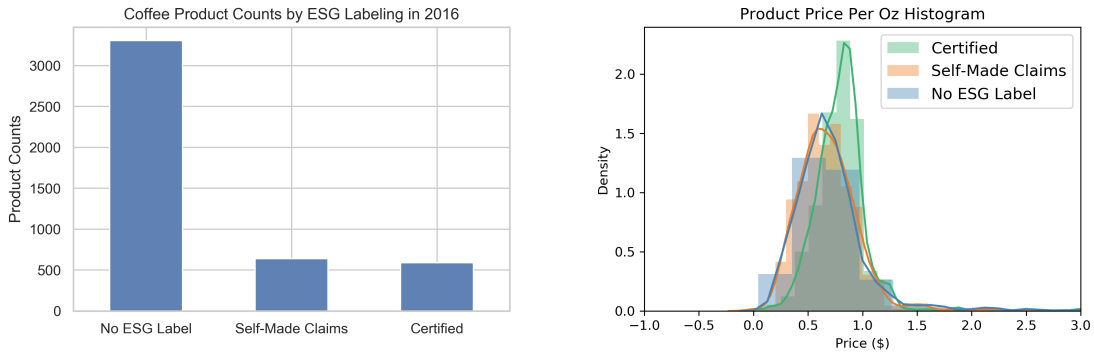


Figure 3 Product counts and prices by category of ESG labeling.

Data summarized from Nielsen Label Insights and Retail Scanner data. Nielsen Label Insights was used to classify products based on ESG marketing on product packaging; the categories are products certified through a major 3rd party certifier (*e.g.*, Fair Trade), products making a reputation-backed ESG claim, or none.

2.3. Control Variables

Appendix B explains our control variables reflecting firm-level production scale, news visibility, and mix of coffee quality level. We utilize these variables to normalize violations rates between companies in empirical regression models which rely on cross-sectional comparisons. We omit them in favor of company fixed effects in regression models which solely utilize within-firm variation to reveal how firms' violation rates respond to their chosen ethical sourcing strategies over time.

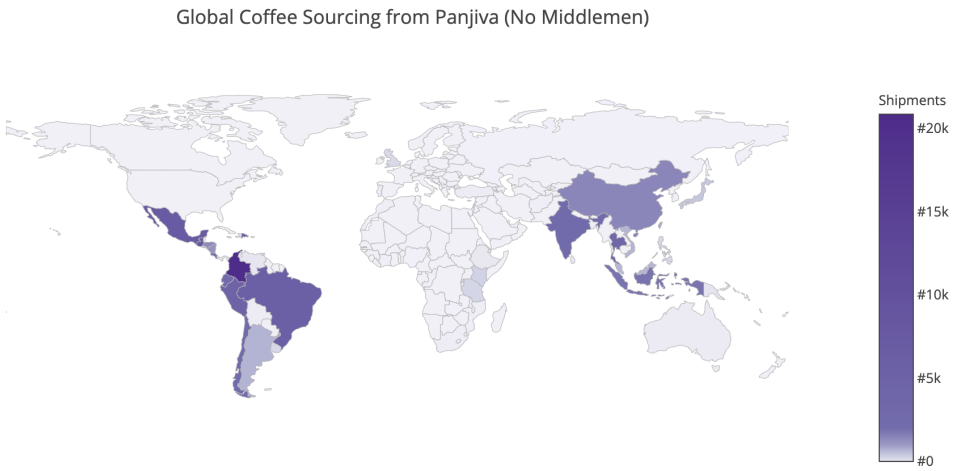
3. Hypotheses

We describe the potential benefits and pitfalls of dual-sourcing and then develop our empirical hypotheses.

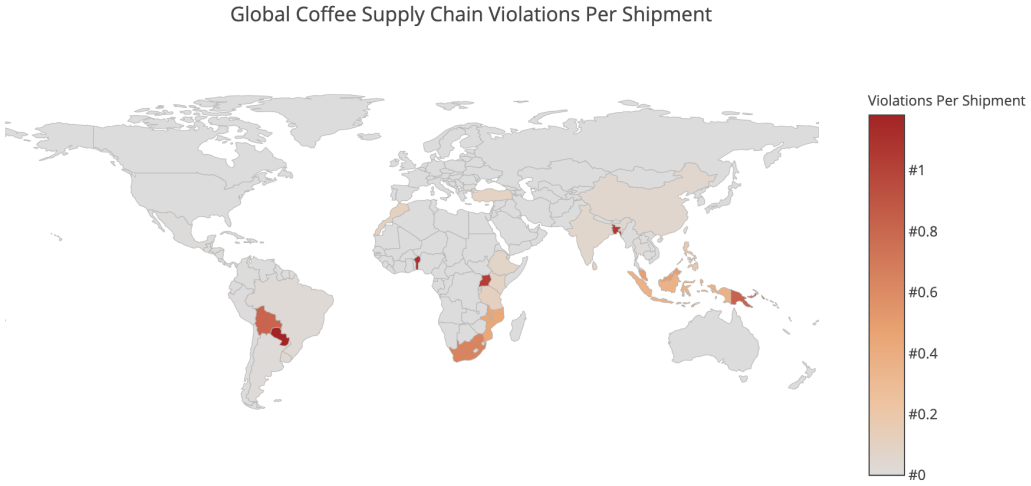
3.1. Stakeholders' Concerns about Dual-Sourcing

The prevalence of dual-sourcing strategies has raised concerns among NGOs, certifiers, and academics (Giovannucci and Koekoek 2003, Trudel and Cotte 2009, Thorlakson et al. 2018).

First, to the extent certification's goal is to achieve industry-wide changes in supplier behavior, dual-sourcing is potentially troubling. By catering to the subset of socially responsible consumers through certified goods while supplying questionably sourced goods to the rest of the market, dual-sourcing may detrimentally segment the market. Empirically, the market shares of certified products often suspiciously echo the shares delivered into the world's wealthiest, most eco-conscious consumer markets. For example, the Roundtable on Sustainable Palm Oil (RSPO) reports certifying 15% of global palm oil in 2012–2013, while US and Europe—the source of nearly all demand for sustainable palm oil—consumes roughly 13–14% of the global supply (Waldman and Kerr 2014). Raising similar concerns, we find that dual-sourcing US coffee companies placed ESG-related claims



(a) Sourcing map of first-tier suppliers, excluding non coffee-producing countries.



(b) Country level ESG violations (extracted from RepRisk) per coffee shipment sent to companies in our study (computed from Panjiva).

Figure 4 Panjiva data tracking origin country for shipments of coffee beans into US ports from 2006-2016.

on only 17% of their uncertified products. Empirical studies have yet to confirm the degree to which dual-sourcing firms' uncertified products harbor poor supplier behavior.

Second, the lack of transparency into dual-sourcing companies' supply chains may exert a negative externality by diluting the reputational credibility of certification labels such as Fair Trade. When a dual-sourcing firm is revealed to have ethical issues in its broader supply base, the reputational fallout may affect consumers' perceptions of the certifier's efficacy. A prominent coffee certification label's lead on chain of custody raised this concern in their conversation with us:

"We don't know what companies are doing to source their non-certified products when they have a mix, it is a significant concern of ours... because it impacts the branding of certification."

Thus, dual-sourcing may represent troubling risks for consumers and certifiers alike.

3.2. Potential Benefits from Dual-Sourcing

On the other hand, both anecdotal case studies and our own first-hand conversations with stakeholders suggest that a firm's certified production can generate positive spillover effects for its non-certified production.

First, certification programs facilitate finding and establishing reliable supplier relationships that are useful beyond certified production. The lead in charge of developing and implementing certification standards for a popular certification program noted to us that:

> *"If a socially responsible retailer wants to source from a specific country, they just have to have the right relationships to know what to do. Often, we [the certifier] can help with that."*

In our first-hand interviews, the supply chain leads of several dual-sourcing coffee retailers noted that they source both certified and non-certified goods from the same suppliers or middlemen because they had learned which ones to trust through their certified production. Anecdotal case studies support positive spillovers from successfully cultivating good supplier relationships (Boström 2015).

Second, piloting a few products through certification may equip coffee companies with transferable operational knowledge regarding responsible sourcing. In a collection of anecdotal evaluations, Barry and Cashore (2012) find that certification acts as a "laboratory for learning and demonstrating best practices," ultimately leading to numerous practices adopted in non-certified sourcing channels. These include better management practices, cleaner technologies, and adherence to worker safety standards (Rueda et al. 2017, Locke et al. 2007b, Boudreau 2020). Once firms have developed the requisite operational and institutional knowledge, they may enjoy diminished marginal costs when further expanding the scope of their responsible sourcing to their other suppliers. Exemplifying how a well-placed intermediary can facilitate knowledge transfer, Li & Fung's Vendor Support Services unit works closely with apparel suppliers to improve their efficiency and ESG compliance based on operational know-how gleaned from the 15,000 suppliers in its network (Lee and Tang 2018).

Thus, in contrast to the concerns about dual-sourcing limiting ethical sourcing, dual-sourcing firms may actually convert their experiences from certifying products into sourcing responsibly for the rest of their supply chains.

3.3. Empirical Questions

To shed light on dual-sourcing's effects in practice, we study the underlying relationship between firms' ethical sourcing decisions, certification levels, and incurred violations. Specifically, companies make ongoing decisions about whom to source from, as well as—to varying degrees—providing

developmental support for and creating visibility and oversight into their suppliers' ethical practices. By strengthening their ethical sourcing decisions, firms become able to certify their products while lowering the rates of violations occurring in their supply chains. In addition to assessing firms to certify products, certifiers can play a more involved role at an earlier stage by supporting and advising companies through the implementation phase during which the firms build their ethical sourcing relationships and programs.

In examining the role of learning-based spillovers, we include two main types of effects. First are direct transfers of knowledge from certifiers to companies. As firms engage with certifiers in their efforts to certify products, they benefit from certifiers' accrued operational experience and know-how. Second, additional knowledge and practices are internally transferable. Upon initial certification, a firm may learn which vendors to work with, how to manage standards, how to refine management practices, etc., and successfully improve responsible sourcing across the supply chain. Under both effects, a firm's knowledge-based ability to implement ethical sourcing is disproportionally well developed by initial product certifications within its product line.

3.3.1. Existing models of CSR violations and sourcing. The literature, which does not consider spillover effects by learning from certifying, commonly predicts or assumes two types of relationships between responsible sourcing decisions and CSR violation rates.

In the scenario where companies are large in comparison to individual producers, as in agricultural supply chains featuring many smallholder farmers, a firm's rate of excess violations scales in proportion with the share of small, uncertified suppliers in its supply base. Each uncertified supplier represents, on average, a heightened risk of violations relative to the baseline risk of a certified supplier. In this case, a firm's overall rate of excess ESG violations decreases linearly as the firm's share of certified sourcing increases. For example, see the left panel of Figure 5, in which, at 60% non-certified production, a company doubles its excess violations rate over 30% non-certified production.

The literature (see, e.g., Guo et al. 2015, Chen and Lee 2016) further considers the scenario in which dual-sourcing firms are small in comparison to one or more large uncertified suppliers. While the large uncertified supplier produces for many firms as buyers, each dual-sourcing company bears a risk of reputational fallout, by association, whenever the supplier's overall production activity generates a reported ESG violation. Under this set of facts, which are less applicable to the coffee and chocolate supply chains, a persistently elevated rate of excess ESG violations attaches to a firm until it completely disassociates itself from the uncertified supplier(s) by certifying its entire product line. See the right-hand panel of Figure 5.

Through these models which do not incorporate spillover effects, the literature has explored ESG violations which elevate upon first engaging non-certified suppliers and then escalate as dual-sourcing companies are further exposed to non-certified sourcing.

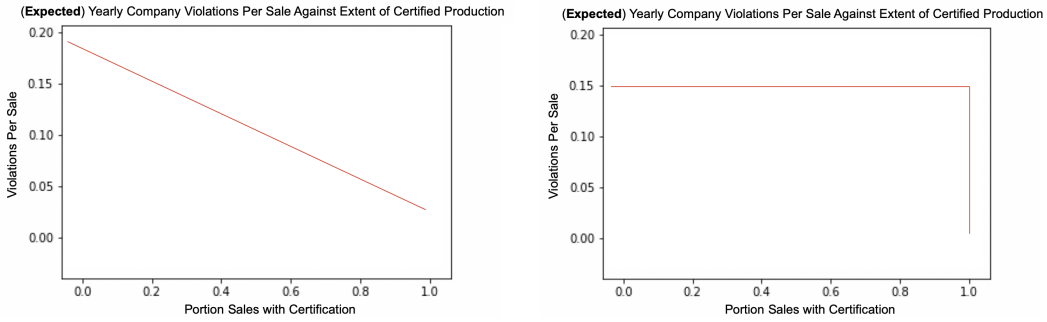


Figure 5 Excess ESG violations per sale against percentage of certified sales under existing models.

The left panel considers ESG risk that scales linearly with the extent of uncertified sourcing; the right panel considers the case where any uncertified sourcing exposes the firm to equal ESG risk.

3.3.2. Learning-based spillovers. Firms' actual yearly ESG violations per sale are plotted against their percentages of certified product sales in Figure 6, with yearly violation time trends removed. In contrast to Figure 5, the empirical pattern reveals a L-shaped relationship under which only firms with very little to no certified sourcing post noticeably higher ESG violation rates. Descriptively, without controlling for company characteristics, ESG violation rates are starkly lower for all companies with any modicum of certified sourcing. This stands in stark contrast to the concerns of NGOs, certifiers, and academics regarding dual-sourcing firms, under which firms which stop certifying their product lines at relatively low thresholds (e.g., 10% or 20%) may greatly under-perform those attaining higher levels of certification. In fact, ESG violation rates appear sharply and similarly low once companies exceed the very modest threshold of 3% certified sales, above which it does not appear consequential whether a firm is 50% or 100% certified, for example. With respect to ESG violations observed in the data, companies which certify just 3% of their sales perform just as well as companies which certify all of their sales.

The L-shape pattern is, however, consistent with learning-based spillovers. In the presence of learning, modest levels of product certification disproportionately improve firms' overall ESG violation rates, since responsibly sourcing a single product lowers the barriers to responsibly sourcing additional products. As certifying firms invest in modifying processes, improving management practices, implementing transparency, and building new relationships to meet the certifier's standards, these learnings and initiatives spill over into the production of other products and encourage responsible sourcing across the supply chain. When spillovers are potent, we expect dual-sourcing firms' excess violations to dissipate quickly after they certify just a few products, as in Figure 6.

Section 4.1's empirical analysis tests for this L-shaped relationship while additionally controlling for company characteristics and, where appropriate, news-based violation detection rates. We also demonstrate a similar L-shaped relationship in the chocolate industry, suggesting that learning-based spillovers are not unique to coffee.

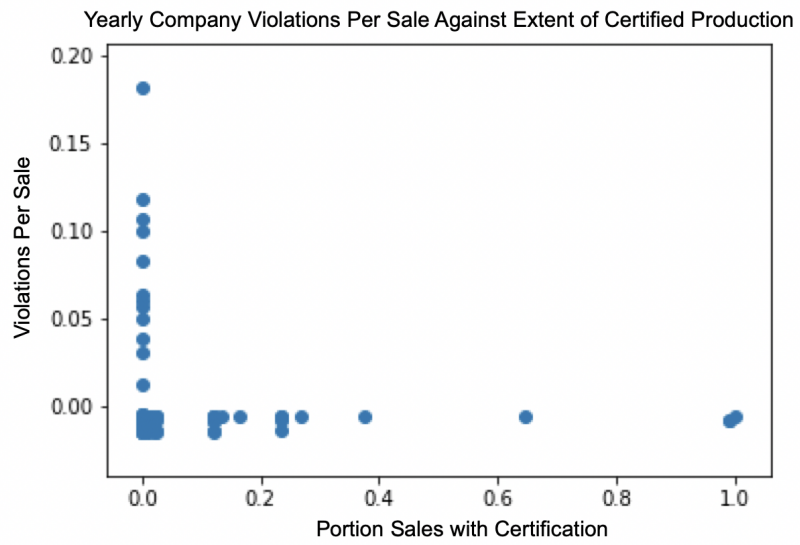


Figure 6 Violations per sale against percentage of sales certified by company-year.

Firms' violations per sale (y-axis) are plotted against their shares of sales from certified products (x-axis) for 47 coffee companies between 2006 and 2016 (398 firm-years). Violations per sale are mean-differenced by year to control for general time trends in violation rates.

3.3.3. Alternative explanations. We identify and address three potential alternative mechanisms that could explain this L-shape relationship. We further discuss why each one is implausible.

First, one could argue that the L-shape is potentially explained by variable time lags between when a firm implements ethical sourcing and formal certification approval. For example, a firm may implement responsible sourcing across its entire supply chain with the intention of certifying broadly, but the certification approval process takes time. Therefore, the reduction in ESG violations will occur contemporaneously with early certifications (following the widespread effort to responsibly source), but only marginal changes will occur with further certification (as administrative approvals are obtained). While this dynamic plausibly explains the L-shape, it is inconsistent with the temporal patterns of firms' certification levels that we observe in the data. Specifically, we would expect to see firms' certification levels increasing over time as their implementation of ethical sourcing is officially verified. However 70% of firms do not have monotonically increasing certification coverage during our 11-year study period, suggesting that this mechanism is unlikely to drive observed firm behavior.

Second, firms that source responsibly may primarily leverage internal sourcing standards and programs to achieve ethical sourcing throughout their supply chain; they may then leverage third-party certifiers to signal their responsible sourcing practices to consumers. Due to the costs associated with certification, some of these firms may opt to certify only a small percentage of their product portfolio while others may opt to certify a larger percentage to amplify signaling to consumers. This

would hypothetically explain the L-shape since, regardless of a firm's signaling strategy, all firms that adopt some level of third-party certification (3% and higher) are already sourcing responsibly. However, such a hypothesis stands in stark contrast to our finding that firms which make their own reputation-backed CSR claims without third-party certification show no improvement in their ESG violation rates. In other words, our results suggest that firms are unable to achieve ethical sourcing based only on internal sourcing standards and programs. One could further hypothesize that firm-specific unobservables distinguish dual-sourcing firms and firms that make reputation-backed claims in terms of their ability to independently operationalize ethical sourcing; however, our use of firm-level fixed effects in a panel study rules out this possibility. In other words, our results strongly suggest that third-party certification improves ESG outcomes for firms (even after controlling for firm-specific unobservables) and does not function as merely a signaling mechanism. This is further supported by our first-hand interviews with stakeholders.

Finally, a third alternative mechanism could be that firms embarking on a dual-sourcing strategy begin by replacing their "worst offenders" (*i.e.*, highest violating suppliers) with certified responsible suppliers first. This would then result in a sharp drop in violations upon initial certification, followed by only marginal returns for replacing the "less offending" remaining suppliers. However, we would then still expect to see gradually diminishing returns from expanding certification coverage, rather than the sharp drop at 3%. For this mechanism to result in our L-shape, we would require firms to have exactly 3% production from high-violating suppliers, with the remaining 97% coming from responsible low-violating suppliers; in this case, 3% certification targeted at the worst suppliers would lead to a drastic reduction in violations, followed by negligible returns since the remaining 97% are already responsible. However, it is highly implausible that all firms have exactly 3% production from high-violating suppliers, particularly given the large diversity in size, country sourcing mix, and initial violation rates among these firms.

3.4. ESG Sourcing by Geography

To understand how certifying affects a firm's ESG performance, we examine the role played by regional sourcing. If certifying firms are turning to countries with well-established production of socially responsible goods, instead of developing and expanding the responsible sourcing base, the gross amount of socially responsible production may not meaningfully change. Thus, the extent to which geographical sourcing mediates certification's ESG improvements importantly indicates whether certifications enable responsibility in difficult conditions, or merely induces firms to source from existing pools of responsible supply.

Thus, we distinguish between ethical sourcing that is accomplished from within "safe" countries versus by undertaking the difficult challenge of sourcing ethically from within high risk regions.

The latter type of ethical sourcing contributes to the broader mission of social responsibility by strengthening emerging economies and creating infrastructure that improves conditions in countries with high supplier risk (Lee and Tang 2018).

Two empirical facts further motivate this analysis. First, US coffee firms' international sourcing decisions vary. Figure 4's top panel shows the heat map of US-destined coffee shipments received from first-tier suppliers located in coffee-producing countries around the globe.

Second, the ESG risk attached to suppliers varies significantly as a geographic function of the source country: some countries are relatively safe (*e.g.*, Peru), while others carry high risk of ESG violations (*e.g.*, Indonesia). Figure 4 contrasts the global heat map of US-destined coffee shipments from coffee-producing countries (top panel) against the countries' ESG violations per coffee shipment (bottom panel).

Underlying such variation, the adoption of certifications and sustainable-sourcing practices by farmers varies by country (Barry and Cashore 2012). On the difficulty of finding reliable suppliers in some parts of the world, the sourcing lead of a regional US-based coffee company with ESG claims told us:

> *"We have a mission to maintain ethical sourcing no matter what country we go to. We tried very hard to source from Indonesia but could not find suppliers that we trusted or mechanisms for making sure our suppliers were maintaining our standards, so we left."*

In Section 4.2's empirical analysis, we account for firms' geographic sourcing decision as a mediating factor in the effect of certifying, or in ESG gains from making independent claims. Additionally, the analysis seeks to validate Section 4.1's examination of the L-shaped relationship by confirming that it holds after controlling for dual-sourcing firms' geographic sourcing.

4. Empirical Analysis

This section details our empirical analyses. In Section 4.1, we empirically test for evidence that dual-sourcing companies realize positive ESG performance spillovers from their certified sourcing into their non-certified sourcing. In Section 4.2, we investigate the role of country sourcing mix in mediating the ESG performance of both certifying firms and those adopting reputation-backed claims, thus potentially explaining the differences in their ESG violation levels.

4.1. Sourcing Spillovers

We examine two empirical claims. We investigate whether certifying more than 3% of sales consistently and substantially lowers a firm's ESG violations. Keeping with the descriptive analysis of Section 3, we further investigate whether additional certifications above this level have no or weak effects on ESG violations. We critically isolate the within-firm variation in certifications and ESG performance, thus removing the cross-sectional correlations across companies. As the dependent

variable measuring firm-year ESG performance, we use $Y_{i,t}$, the yearly count of company-specific ESG violation incidents captured by RepRisk. Section 4.1.1 first analyzes the threshold effect of certifying more than 3% of sales, and then Section 4.1.2 separately analyzes the impact of incrementally increasing certification coverage beyond 3%.

4.1.1. Initial certifications. Social and sustainability focused certification programs operate by imposing and auditing strict standards on working conditions, wages, and environmental impacts and revoking certification for companies that repeatedly fail to match certification standards. Thus, participation is expected to lower ESG violation incidents.

Companies that certify more than 3% of their production are expected to have lower violation rates due to their compliance with certification standards. When considering the treatment, $\tau_{i,t}$, an indicator variable that represents whether a company has more than 3% of their production certified in a particular year, we test this hypothesis under the following specifications:

$$Y_{i,t} = \alpha\tau_{i,t} + \xi_i + \gamma X_{i,t} + T_t + \epsilon_{i,t}.$$

where ξ_i are company fixed effects, $X_{i,t}$ are company controls, and T_t are time fixed effects. We include company fixed effects to analyze the within-company variation in violation incidents in response to the firm's changes in certification coverage, including after passing the 3% initial certification threshold.

Results: The regression results for the coffee industry for 47 firms and 398 firm-years are given in Table 2. We are interested in the effect size of our treatment variable, $\tau_{i,t}$, an indicator variable that represents whether a company has more than 3% of their production certified in a particular year, on our outcome variable, ESG violations. As expected, the results show that the treatment of having more than 3% certified sales is related to lower ESG violations with a significant effect size of -14.9 violations per year. This is a fairly sizable reduction in yearly violations given that companies with violations experience an average of 23 violations per year. We first test these results using controls for company fixed effects, production volume and news visibility and then add controls for time trends. All control coefficients represent effect sizes for variation within a company because we are incorporating company fixed effects.

We verify our results by applying the same specifications to the chocolate industry for 100 firms and 944 firm-years to confirm that violations similarly fall when a firm certifies more than 3% of its production (see Table 3). We exclude the control of news count from the chocolate analysis due to changes in the availability of data. We find a significant treatment effect of -112.5 violations per year for companies certifying more than 3% of their portfolios. Chocolate companies have higher violation rates, with companies with any violations averaging 168 violations per year. Therefore,

Outcome: Yearly ESG violations by coffee company

Variable	(1) Certification Threshold		(2) + Year FE	
	Estimate	**Std Error**	**Estimate**	**Std Error**
3%+ Certified Sales	-11.9^{**}	(4.23)	-14.9^{***}	(4.14)
Log(Sales)	-0.13	(0.62)	-0.91	(0.62)
News Count	-0.015^{*}	(0.007)	-0.017^{*}	(0.007)
Company Fixed Effect	✓		✓	
Year Fixed Effect			✓	
r^2	0.51		0.55	

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 2 Regression results for ESG violations by sourcing segment for coffee companies

Regression estimates using a panel of 398 firm-years.

the results from the chocolate industry further corroborate our results from the coffee industry that certifying more than 3% of production has on positive impact on social and environmental standards.

Outcome: Yearly ESG violations by chocolate company

Variable	(1) Certification Threshold		(2) + Year FE	
	Estimate	**Std Error**	**Estimate**	**Std Error**
3%+ Certified Sales	-101.8^{***}	(20.1)	-112.5^{***}	(20.2)
Log(Sales)	-1.49	(1.8)	1.8	(1.8)
Company Fixed Effect	✓		✓	
Year Fixed Effect			✓	
r^2	0.93		0.93	

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 3 Regression results for ESG violations by sourcing segment for chocolate companies

Regression estimates using a panel of 944 firm-years.

4.1.2. Certifications above 3%. Next, we consider the treatment variable of percent certified sales, focusing on the segment of companies certifying more than 3% of their sales. This serves to test the effect of incremental further certifications on ESG violations for the subset of companies that already have 3%+ certification rates. We apply the following specification:

$$Y_{i,t} = \alpha_1 \delta_{i,t} + \xi_i + \gamma X_{i,t} + T_t + \epsilon_{i,t}.$$

where the covariate $\delta_{i,t}$ represents the percent of certified sales for a particular company-year.

We can then compare the effect size of passing the 3% certification threshold, $\tau_{i,t}$ from Section 4.1.1, with the effect of coverage expansions for the regime of companies already sourcing more than 3% from certified channels, $\delta_{i,t}$.

Due to the smaller sample size, we combine companies from both the coffee and chocolate industries. To corroborate strong learning effects, we expect to see no significant effect from the degree of certification, $\delta_{i,t}$, implying that increasing certification coverage levels, once already certified, does not lead to further gains in lowering ESG violations.

Results: We find that the actual amount of a company's portfolio that is certified (Percent Sales Certified) beyond 3% of sales does not significantly impact ESG violations (see Table 4). Therefore, it is having more than 3% certified production which drives the knowledge-based spillovers throughout the overall supply chain including the non-certified portions.

The estimated effect, $\delta_{i,t}$, is also quite economically small. For example, we find that moving firms with modest certification levels above the 3%+ threshold results in a significant reduction in ESG violations for the studied companies of 6.5%. The equivalent increase in certification for companies that are already 3%+ certified (derived from the value of $\delta_{i,t}$) only reduces violations by 0.3 per company-year, amounting to a 0.1% reduction in overall ESG violations for the coffee companies studied. The former amounts to a 46× greater impact on reducing violations outcomes. For a company with 3%+ certified sales to have the equivalent impact on reducing violations as a firm moving minimally across the 3% threshold, it would need to expand certification levels by 90%.

Outcome: Yearly ESG violations by company (chocolate and coffee)

Variable	**Estimate**	**Std Error**
Percent Sales Certified	−15.9	(12.1)
Log(Sales)	−1.02	(1.4)
Company Fixed Effect	✓	
Year Fixed Effect	✓	
r^2	0.95	

$^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 4 Regression results for yearly ESG violations for chocolate and coffee companies with more than 3% certified sales for 104 firm-years.

4.2. Certification Effects Mediated by Geographical Sourcing

We examine the role played by country-based sourcing in certification's effects on ESG performance. Certifying firms' geographical sourcing decisions can mediate the reductions they achieve in ESG violations. For example, a firm may certify by choosing low-violation suppliers located in *low risk* countries that enjoy robust monitoring and enforcement. The firm sources ethically by leveraging an established pool of responsible suppliers. In contrast, a certifying firm may source ethically from within *high risk* countries in which corruption and ESG violations are endemic. When firms responsibly source from such countries, it positively impacts the region and expands

the amount of local socially responsible production (Lee and Tang 2018). Therefore, the extent to which geographical sourcing mediates certification's effects on ESG performance reflects whether certifying firms merely switch to existing responsible supplier pools or cultivate new responsible sourcing options in challenging regions.

First, we simply examine violations as a function of company segment without accounting for the mediating effect of country sourcing. As our response variable, $Y_{i,t}$, we use the yearly count of detected ESG violations for a firm i in year t from RepRisk. We add covariates $\mathbb{1}_{>3\%\,\text{Cert,i}}$ and $\mathbb{1}_{\text{Claims,i}}$ as indicator variables that denote if firm i has 3+% certified sales and is making reputation-backed ethical sourcing claims respectively. We add yearly fixed effects, T_t, to control for violation/detection trends over time, and product sales volume, $\log(\text{Sales})_{i,t}$, to control for company size. We investigate the relationship between company segment and violation outcomes through the following specifications:

$$Y_{i,t} = \alpha_0 \mathbb{1}_{>3\%\,\text{Cert,i,t}} + \alpha_1 \mathbb{1}_{\text{Claims,i,t}} + T_t + \alpha_3 \log(\text{Sales})_{i,t} + \varepsilon_{i,t}.$$

Second, we are interested in removing the mediating factor of firms' geographical sourcing strategies that could explain, in part or in whole, the effect of certification on improving ESG performance. Therefore, we now add controls for the share of sourcing from each of the top 17 coffee-producing countries (these countries alone make up 95% of global coffee production (International Coffee Organization 2019)). Including controls for firms' regional sourcing allows us to separate the role sourcing strategies play at reducing ESG violations and isolate the impact of participating in certification programs. Furthermore, we can also compare the role of country sourcing strategies on ESG performance for dual-sourcing companies against companies with self-made claims. Therefore, we use the following specifications:

$$\begin{aligned} Y_{i,t} = {} & \alpha_0 \mathbb{1}_{>3\%\,\text{Cert,i,t}} + \alpha_1 \mathbb{1}_{\text{Claims,i,t}} + T_t + \alpha_3 \log(\text{Sales})_{i,t} \\ & + \sum_{c=1}^{17} \beta_c^B (\text{Coffee Producing Country})_{i,c} + \varepsilon_{i,t}. \end{aligned}$$

where, $(\text{Coffee Producing Country})_{i,c}$ denotes the fraction of firm i's coffee shipments that arrive from country c based on our Panjiva data.

Results: We provide the results for the coffee industry in Table 5. In the first piece of our analysis (regression (1) in Table 5), we are interested in comparing the effect of belonging to each company segment: (1) dual-sourcing with more than 3% certified sales, (2) companies with self-made reputation backed claims, and (3) control companies with no CSR sourcing initiatives. Unsurprisingly, we see that dual-sourcing companies have lower violations than companies with no ESG policies, consistent with our results in Section 4.1.1. Furthermore, dual-sourcing companies

maintain this advantage over companies with reputation-backed claims made from self-led ESG programs. Table 5 shows that these companies with self-made claims do not have significantly lower ESG violations than the control companies that don't advertise any claims around ESG policies.

The second piece of our analysis (regression (2) in Table 5) controls for country sourcing strategies in order to remove it as a mediating factor in reducing ESG violations. We see that after removing the mediating factor of geographical sourcing, certifying firms have 10.9 fewer annual violations relative to the control group, a significant jump from regression (1) without country controls. These certifying companies source more frequently from high risk countries than their peers, with 45% of their sourcing coming from high risk regions (in contrast to 34% for companies with self-made claims and 29% for remaining companies).[2] Therefore, once we control for their sourcing from high risk locations, we see that the advantage dual-sourcing companies have at reducing violations increases significantly. Therefore, not including controls for the mediating factor of country sourcing would lead to a sizeable underestimation in treatment effect of certifying more than 3% of sales.

Furthermore, this highlights that dual-sourcing companies are able to maintain low violations *despite* sourcing from high risk regions. In other words, dual-sourcing companies appear to be creating infrastructure and processes that support responsible sourcing even when operating in locations where it is extremely challenging to do so (e.g. locations identified in Figure 4).

In contrast, controlling for sourcing mechanisms does not significantly impact the effect size for the company segment with self-made claims. Therefore, the lack of ESG violation reduction in these companies is not due to sourcing from high risk locations where it is challenging to maintain ethical sourcing. Rather, these companies are not implementing programs effectively across their supply chains to improve responsible sourcing.

5. Discussion & Concluding Remarks

This paper highlights a novel role of certification in enabling companies to overcome the initial barriers of responsible sourcing, rather than a blanket solution intended to cover entire supply chains. Our primary finding is that certifying firms' rates of CSR violations are similarly low regardless of whether the firm's portfolio is 5% certified or 100% certified. In other words, taking a *first step* towards responsible sourcing by certifying a small portion of the product portfolio is a key first-order challenge that enables learning-based spillovers and facilitates responsible sourcing expansion. Our secondary finding is that certifying companies are uniquely able to maintain low violations despite their prevalent sourcing from "high risk" countries, allowing them to expand the responsible supplier pool in these regions.

[2] We define "high risk" countries as the top ten countries with the highest ESG violations accumulated across all industries according to the RepRisk data used in this study. We find consistent results across alternative definitions of "high risk" countries, *e.g.*, countries with high ESG violations per capita, or limiting to coffee-specific ESG violations.

Outcome: Yearly ESG violations by company (coffee)

Variable	(1) Without Sourcing		(2) With Sourcing	
	Estimate	Std Error	Estimate	Std Error
3%+ Certified	−4.66	3.2	−10.9***	3.1
Self-made claims	−1.1	2.6	−2.96	2.67
Company Controls	✓		✓	
Year Fixed Effects	✓		✓	
Country Sourcing Controls			✓	
r^2	0.08		0.37	

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 5 Regression results for yearly ESG violation outcomes for coffee company segments when controlling for country sourcing mechanisms for 372 firm-years. Please note that 4 firms with no available Panjiva data were dropped from the sample for this analysis.

Our findings motivate our view of certification-based responsible sourcing as a relatively *binary* outcome (we do not observe partially certified firms with "moderate" violations). Our data suggests that it is not necessary for firms to invest in 100% certified product portfolios to achieve responsible sourcing (in contrast to, *e.g.*, Guo et al. 2015, Chen and Lee 2016). This finding is important since our primary interviews with leaders of certified coffee supply chains suggest that certifying the entire product portfolio is an onerous task, and may actually hurt responsible sourcing outcomes by limiting entry into some emerging economies. Current certification standards are typically applied globally; however, it is well-recognized that differences in local culture and labor markets can make inflexible standards difficult to adopt in certain regions (Barry and Cashore 2012, Boström 2015). For example, the supply chain lead at a UK coffee company told us,

> *"We need exceptions based on specific country conditions, for example in India the agricultural labor standard in most certifications would require law changes that would either be extremely expensive and/or difficult to change. In addition, housing and living conditions in the standards do not account for cultural differences in different countries, for example housing capacity limits [intended to prevent unsafe or exploitative crowding in worker living conditions] are too small for Indian families that want to live together."*

As discussed in Section 3.4, another sourcing lead described challenges in finding responsible suppliers in "risky" countries such as Indonesia. The lack of flexibility in current certification standards greatly limits the ability for socially-conscious firms to source from emerging economies in their certified production. Thus, given that a primary goal of responsible sourcing is to support welfare improvements in emerging economies—*e.g.*, sourcing by responsible firms has been shown to improve incomes, education rates, and regulatory governance (Dragusanu and Nunn 2018)—pushing for blanket certification of supply chains may *hinder* rather than support this goal.

Fortunately, our country-based analysis suggests that dual sourcing firms are able to leverage learning-based spillovers to responsibly source from "risky" countries despite the lack of infrastructure for doing so. This result motivates the view of certifications as "learning labs," allowing firms to learn how to create and manage reliable supplier relationships while circumventing the inflexibility of current certification standards. This is an important contribution on the part of dual-sourcing firms both in the short term (achieving low ESG violations while improving welfare in risky countries), and the long term (in many cases, local governments have adopted voluntary standards set by these firms (Barry and Cashore 2012)). Thus, a promising agenda for future research is designing appropriate incentives that encourage dual-sourcing firms to engage emerging economies and create the requisite infrastructure for ethical sourcing.

Acknowledgments

The authors gratefully acknowledge funding from the Wharton Social Impact Initiative and research assistance from Angel Chung and Avideep Pradhan. We are indebted to Vishal Agrawal, Saed Alizamir, Noah Gans, Vishal Gaur, Amrita Kundu, Somya Singhvi, Karen Zheng, and various seminar participants for providing valuable feedback on earlier drafts.

Our analyses are derived based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of the researcher(s) and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

References

Abbasi, Maisam. 2017. Towards socially sustainable supply chains – themes and challenges. *Eur. Business Rev.* **29**(3) 261–303.

Akkaya, Duygu, Kostas Bimpikis, Hau Lee. 2020. Government interventions to promote agricultural innovation. *Manufacturing & Service Oper. Management* .

Amengual, Matthew, Greg Distelhorst. 2019. Can sourcing help enforce global labor standards? Evidence from the Gap Inc supply chain. *Evidence from the Gap Inc Supply Chain (October 9)* .

Babich, Volodymyr, Christopher S Tang. 2012. Managing opportunistic supplier product adulteration: Deferred payments, inspection, and combined mechanisms. *Manufacturing & Service Oper. Management* **14**(2) 301–314.

Barry, Mike, Ben Cashore. 2012. The roles and limitations of certification .

Bastani, Hamsa, Joann F de Zegher. 2019. Do policies with limited enforcement reduce harm? Evidence from transshipment bans. *SSRN Working Paper* (Add number).

Bloom, Nicholas, Christos Genakos, Ralf Martin, Raffaella Sadun. 2010. Modern management: Good for the environment or just hot air? *Econom. J.* **120**(544) 551–572.

Boström, Magnus. 2015. Between monitoring and trust: commitment to extended upstream responsibility. *Journal of business ethics* **131**(1) 239–255.

Boudreau, Laura. 2020. Multinational enforcement of labor law: Experimental evidence from Bangladesh's apparel sector. *Working paper* .

Buell, Ryan W, Shwetha Mariadassou, Yanchong Zheng. 2019. Relative performance transparency: Effects on sustainable choices. *Harvard Business School Technology & Operations Mgt. Unit Working Paper* (19-079).

Caro, Felipe, Leonard Lane, Anna Sáez de Tejada Cuenca. 2018. Can brands claim ignorance? Unauthorized subcontracting in apparel supply chains. *UCLA Anderson working paper* .

Castka, Pavel, Charles J Corbett, et al. 2015. Management systems standards: diffusion, impact and governance of ISO 9000, ISO 14000, and other management standards. *Foundations and Trends® in Technology, Information and Operations Management* **7**(3–4) 161–379.

Chen, Jiayu, Anyan Qi, Milind Dawande. 2019. Supplier centrality and auditing priority in socially-responsible supply chains. *Manufacturing & Service Oper. Management* (forthcoming).

Chen, Li, Hau L Lee. 2016. Sourcing under supplier responsibility risk: The effects of certification, audit, and contingency payment. *Management Sci.* **63**(9) 2795–2812.

Chen, Shi, Qinqin Zhang, Yong-Pin Zhou. 2018. Impact of supply chain transparency on sustainability under NGO scrutiny. *Prod. Oper. Management* .

Cho, Soo-Haeng, Xin Fang, Sridhar Tayur, Ying Xu. 2019. Combating child labor: Incentives and information disclosure in global supply chains. *Manufacturing & Service Oper. Management* .

Consortium, The Sustainability. 2016. Greening global supply chains: From blindspots to actions). URL `https://www.sustainabilityconsortium.org/wp-content/themes/enfold-child/assets/pdf/2016-impact-report.pdf`.

Corbett, Charles J, María J Montes-Sancho, David A Kirsch. 2005. The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management Sci.* **51**(7) 1046–1059.

Darnall, Nicole, Stephen Sides. 2008. Assessing the performance of voluntary environmental programs: Does certification matter? *Policy Stud. J.* **36**(1) 95–117.

de Zegher, Joann F, Dan A Iancu, Erica Plambeck. 2018. Sustaining rainforests and smallholders by eliminating payment delay in a commodity supply chain–it takes a village. *Working paper* .

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

Dragusanu, Raluca, Nathan Nunn. 2018. The effects of Fair Trade certification: Evidence from coffee producers in Costa Rica. Tech. rep., National Bureau of Economic Research.

Fang, Xin, Soo-Haeng Cho. 2019. Cooperative approaches to managing social responsibility in a market with externalities. *Manufacturing & Service Oper. Management* (Forthcoming).

Giovannucci, Daniele, Alice Byers, Pascal Liu. 2008. Adding value: Certified coffee trade in north america .

Giovannucci, Daniele, Freek Jan Koekoek. 2003. The state of sustainable coffee: A study of twelve major markets .

Guo, Ruixue, Hau L Lee, Robert Swinney. 2015. Responsible sourcing in supply chains. *Management Sci.* **62**(9) 2722–2744.

Guo, Xiaomeng, Guang Xiao, Fuqiang Zhang. 2017. Effect of consumer awareness on corporate social responsibility under asymmetric information. *Available at SSRN 3039862* .

Hasle, Peter, Per Langaa Jensen. 2012. Ergonomics and sustainability–challenges from global supply chains. *Work* **41**(Supplement 1) 3906–3913.

Huang, Lu, Jing-Sheng Jeannette Song, Robert Swinney. 2017. Managing social responsibility in multitier supply chains. *Available at SSRN 2837332* .

ILO. 2018. Human rights in supply chains (human rights watch). URL `https://www.dol.gov/agencies/ilab/our-work/child-forced-labor-trafficking/child-labor-cocoa`.

International Coffee Organization. 2019. International coffee organization: Trade statistics tables .

Kalkanci, Basak, Erica L Plambeck. 2018a. Managing supplier social & environmental impacts with voluntary vs. mandatory disclosure to investors. *Georgia Tech Scheller College of Business Research Paper* (37) 16–5.

Kalkanci, Basak, Erica L Plambeck. 2018b. Reveal the supplier list? a trade-off in capacity vs. responsibility. *Georgia Tech Scheller College of Business Research Paper* (18-4) 18–8.

Kim, Sang-Hyun. 2015. Time to come clean? Disclosure and inspection policies for green production. *Oper. Res.* **63**(1) 1–20.

Kraft, Tim, León Valdés, Yanchong Zheng. 2018. Supply chain visibility and social responsibility: Investigating consumers' behaviors and motives. *Manufacturing & Service Operations Management* **20**(4) 617–636.

Kraft, Tim, León Valdés, Yanchong Zheng. 2019. Motivating supplier social responsibility under incomplete visibility. *Manufacturing & Service Oper. Management* (forthcoming).

Lee, Hau L, Christopher S Tang. 2018. Socially and environmentally responsible value chain innovations: New operations management research opportunities. *Management Sci.* **64**(3) 983–996.

Levi, Retsef, Manoj Rajan, Somya Singhvi, Yanchong Zheng. 2019. Unifying agricultural wholesale markets: Impact on farmers' income. *Available at SSRN* (Add number).

Lewis, Tracy R, Fang Liu, Jing-Sheng Song. 2019. A dynamic mechanism for achieving sustainable quality supply. *Fuqua School of Business, Duke University, Durham* .

Li, Jun, Di Andrew Wu. 2017. Does corporate social responsibility benefit society? *Management Sci.* (forthcoming).

Locke, Richard, Thomas Kochan, Monica Romis, Fei Qin. 2007a. Beyond corporate codes of conduct: Work organization and labour standards at Nike's suppliers. *Internat. Labour Rev.* **146**(1-2) 21–40.

Locke, Richard M, Fei Qin, Alberto Brause. 2007b. Does monitoring improve labor standards? Lessons from Nike. *ILR Rev.* **61**(1) 3–31.

Nelson, Valerie, Ximena Rueda, Walter JV Vermeulen. 2018. Challenges and opportunities for the sustainability transition in global trade (introduction). *Business Strategy and the Environment* **27** 173–178.

Orsdemir, Adem, Bin Hu, Vinayak Deshpande. 2019. Ensuring corporate social and environmental responsibility through vertical integration and horizontal sourcing. *Manufacturing & Service Oper. Management* **21**(2) 417–434.

Pagell, Mark, Zhaohui Wu. 2009. Building a more complete theory of sustainable supply chain management using case studies of 10 exemplars. *J. Supply Chain Management* **45**(2) 37–56.

Plambeck, Erica L, Terry A Taylor. 2015. Supplier evasion of a buyer's audit: Implications for motivating supplier social and environmental responsibility. *Manufacturing & Service Oper. Management* **18**(2) 184–197.

Porteous, Angharad H, Sonali V Rammohan, Hau L Lee. 2015. Carrots or sticks? Improving social and environmental compliance at suppliers through incentives and penalties. *Prod. Oper. Management* **24**(9) 1402–1413.

Rueda, Ximena, Rachael D Garrett, Eric F Lambin. 2017. Corporate investments in supply chain sustainability: Selecting instruments in the agri-food industry. *J. Cleaner Prod.* **142** 2480–2492.

Sodhi, ManMohan S, Christopher S Tang. 2018. Research opportunities in supply chain transparency .

The Guardian. 2011. How green is your coffee? URL `https://www.theguardian.com/environment/2011/oct/04/green-coffee`.

Thorlakson, Tannis, Joann F de Zegher, Eric F Lambin. 2018. Companies' contribution to sustainability through global supply chains. *Proc. National Acad. Sci.* **115**(9) 2072–2077.

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

Trudel, Remi, June Cotte. 2009. Does it pay to be good? *MIT Sloan Management Rev.* **50**(2) 61.

US Department of Labor. 2013. US Labor Department's initiative results in more than $100,000 in back wages and penalties for violations affecting 150 workers in Hawaii's coffee industry. URL `https://www.dol.gov/whd/media/press/whdpressVB3.asp?pressdoc=Western/20130221.xml`.

Waldman, Kurt B, John M Kerr. 2014. Limitations of certification and supply chain standards for environmental protection in commodity crop production .

WWF. 2007. Gone in an instant URL `https://wwf.panda.org/wwf_news/?92080/`.

Zamora, Miguel. 2013. Farmworkers left behind: The human cost of coffee production. *Daily Coffee News* URL `https://dailycoffeenews.com/2013/07/17/farmworkers-left-behind-the-human-cost-of-coffee-production/`.

Zhang, Han, Goker Aydin, Rodney P Parker. 2019. Social responsibility auditing in supply chain networks .

Appendix

A. Certification Labels

Table 6 and Table 7 include the certification endorsements on coffee and chocolate products captured by the Nielsen LabelInsights data in 2016. On-pack certification labels were used to identify the percentage of products in a firm's portfolio that were produced under certified standards. Out of 4,534 coffee products, 13% had formal certification endorsements on their packaging. Within the group of certified coffee products, 71% of sales were Fair Trade, 35% were Rainforest Alliance and 9.6% were other programs (please note that a few products had multiple certifications, therefore will belong in multiple buckets).

Coffee Certifications
Certified B Corporation
Certified SC Product
Federación Nacional de Cafeteros
Cooperative Coffees Member
Fair Trade Certified
Fair trade certified by Pareve
Fair Trade USA
Rainforest Alliance Certified
100% Rainforest Alliance Certified
30% Rainforest Alliance Certified
65% Rainforest Alliance Certified
Social Certified by IMO

Table 6 Sustainability-related Third-party Certifications for Consumer Coffee Products

Chocolate Certifications
100% Fair Trade Certified
Certified B Corporation
Certified Sustainable Palm Oil
Cocoa Life Certified
Direct Trade Certified
Fair for Life Certified
Fair Trade Certified
Fair Trade USA Certified
Fairwild Certified
Green Energy Certified
Heirloom Cacao Preservation Certified
Project Seal
Rainforest Alliance Certified
True Source Certified
Worker Owned Coop

Table 7 Sustainability-related Third-party Certifications for Consumer Chocolate Products

B. Control Variables

We leverage a series of control variables reflecting firm-level production scale, news visibility, and mix of coffee quality level. We utilize these variables to normalize violations rates between firms in empirical regression models which rely on cross-sectional comparisons. We omit them in favor of firm fixed effects in regression models which solely utilize within-firm variation to connect firms' violation rates to their chosen ethical sourcing strategies over time.

B.1. Scale of production.

To appropriately interpret firms' ESG violation counts, we control for company-level production scale. Thus, we account for the fact that, *e.g.*, companies producing at a larger scale accumulate higher violation counts, even when keeping their violations *per sale* low. Therefore, we collect product-level sales data from Nielsen Retail Scanner and then aggregate to obtain company-level sales and revenue.

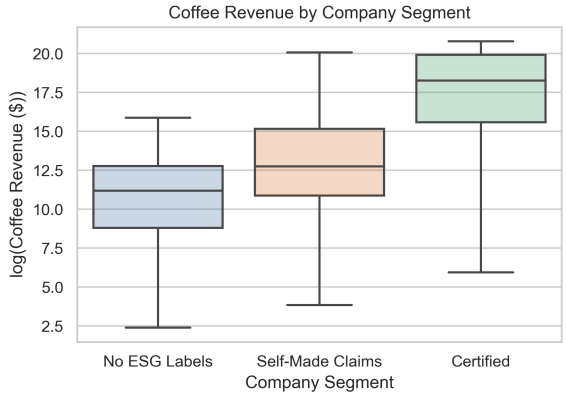


Figure 7 Boxplot of company-level coffee revenue by category of ESG labeling.

Data from Nielsen Label Insights and Retail Scanner data from 2006-2016. Each data point is a company's coffee revenue in a given year. Companies are categorized based on having 3%+ of their product sales with third-party certification, remaining companies with some reputation-backed claims on ESG sourcing, and companies with no ESG claims.

Figure 7 plots companies' coffee revenues in each of the three company categories. Companies larger by sales and revenues tend to have at least one product line with an ESG claim, whether third-party certified or reputation-backed.

B.2. News visibility.

RepRisk compiles firms' violation events when they are detected by news and regulatory sources, and the rates of detection likely vary with firms' news exposure levels. Thus, among a pair of companies incurring hypothetically the same violations, the company subject to higher news volume and scrutiny likely sees a greater number of its incidents brought to light and tallied. To measure a company's level of news exposure, we compute and use the volume of general news published about the company during the year of interest collected from LexisNexis.

B.3. Specialty/commodity share of production.

Coffee companies have different product portfolio mixes of specialty versus commodity goods. The supply chains for companies that focus on specialty goods can look different than the supply chains of companies that primarily sell commodity coffee. To control for this choice in our robustness checks, we add controls of the average price in a company's product portfolio. Companies offering primarily specialty goods tend to have a higher average price across their portfolio.

C. Dataset Matching Summary

The goal of our study is to analyze an overall firm's portfolio of products to understand the impact of firm level CSR strategies on responsible sourcing. Therefore, we merged multiple datasets to provide visibility into product portfolios, sales volume, violations, and sourcing for individual firms. Merging datasets requires matching firms that exist in both datasets. Thus, our final sample after merging will contain only firms that exist in all datasets. Table 8 shows the resulting sample size upon merging each additional dataset.

Our final dataset covers 59% of the entire US coffee market and 93% of the certified coffee sales in the US. Since our goal is to understand how companies with certifications perform across varying levels of certified coverage, we have high coverage of the relevant portions of the coffee industry for our study. Many of the uncovered coffee sales arise from firms that do not have visibility or sales in retail outlets (and are therefore not represented in our Nielsen LabelInsights dataset); even if we had data from these companies, they would not provide an effective control group against the certified firms we study, which tend to have high visibility and sales at retail outlets. Finally, to boost the credibility of our findings, we have invested significant effort to collect an analogous dataset from the chocolate industry; we find reassuringly similar insights.

Data Source	Share of Total Sales	Share of Certified Sales	Brands	Companies	Products
Retail Scanner	100%	-	-	-	7,591
+ LabelInsights	63%	100%	527	-	4,534
+RepRisk	59%	93%	130	47	1,607
+Panjiva	58%	92%	126	43	1,554

Table 8 **Data filtering results for the Coffee Industry after matching to each additional dataset.**

D. Detection Bias Simulation

There is unobserved variation in detection of violations across countries. For example, some high violating countries have constrained resources and/or higher corruption that reduces their ability to report on true domestic violation rates. This leads to concern that detection bias could be driving the L-shape results we uncover (e.g. if dual-sourcing companies are sourcing primarily from countries with poor detection, their observed violations would be much lower than their true value, potentially resulting in an empirical L-shape relationship between violations and certification coverage). In this section, we simulate varying levels of violation detection rates across countries and find that detection bias is not a feasible alternative mechanism behind the L-shape relationship identified.

D.1. Simulation Setup

We now describe our simulation setup, which we use to explore the impact of detection bias on the observed relationship between certification coverage and detected violations.

We include the 118 countries that our coffee companies source from in the simulation. Each country is assigned to be either a *high risk* or *low risk* country based on their violation rates extracted from RepRisk. The countries in the top quartile of violations are considered *high risk* countries, while the remaining are considered *low risk*. Violations are simulated for each country for 10 time periods (years) and are generated using the normal distribution where μ and σ are assigned based on the RepRisk data for each country segment. For example, simulated violations for *high risk* countries are generated from the normal distribution with μ_{high} and σ_{high}, which are pulled directly from the RepRisk data on ESG violations in these *high risk* countries for coffee companies. *High risk* countries are assigned a detection bias, which determines what percentage of the country's violations are actually detected (e.g. a detection bias of 10% signifies that 10% of violations are not detected, thus the reported violations are only 90% of the actual violation counts). We simulate detection bias levels from 0-100% in increments of 10%.

We simulate the 43 coffee companies in which we have both sourcing data from Panjiva and violation data from RepRisk. Each company is assigned sourcing levels from each of the 118 countries based on the sourcing connections identified from Panjiva. Therefore, the number of countries sourced from and the mix of *high risk* and *low risk* countries for each company is taken directly from the Panjiva data. In addition, each company is assigned a level of certification coverage based on their data from Nielsen LabelInsights.

Each company is assigned detected ESG violation counts based the violations generated by the countries they are sourcing from. We assumed violations are linearly decreasing in certification coverage to test whether detection bias can distort a true underlying linear pattern into an L-shape relationship between certification coverage and violations. We run 1000 simulations of 10 time periods each and compute the detected violations for each company-year.

Next, we evaluate the statistical significance of an L-shape relationship between certification coverage and violations to determine if detection bias is a feasible alternative mechanism. For each simulation run for 43 companies and 10 time periods, we run the same regression specifications from Section 4, first testing whether companies above 3% certification have significantly fewer violations across the entire population:

$$Y_{i,t} = \alpha \tau_{i,t} + \epsilon_{i,t}.$$

where $Y_{i,t}$ is ESG violation counts by company-year and $\tau_{i,t}$ our treatment of whether a company has more than 3% certification coverage. Note, we do not need to include company or time fixed effects because we are not simulating the dynamics behind those covariates. Then, we focus on companies with more than 3% certification to estimate the incremental value of increases in certification coverage beyond 3%:

$$Y_{i,t} = \alpha_1 \delta_{i,t} + \epsilon_{i,t}.$$

where $\delta_{i,t}$ is the continuous value of certification coverage. We use the regression results from these two specifications to determine whether or not that particular simulation resulted in an L-shape relationship between violations and certification coverage.

D.2. Simulation Results

We test detection bias levels in high risk countries of 0-100% in increments of 10%. We find that there is no level of detection bias that can feasibly result in an L-shape pattern between certification coverage and violations. While there is a small range where 70-80% of violations go undetected in high risk countries that can produce an L-shape, this level of detection bias causes high violating countries to have fewer detected violations than low violating countries, which is empirically not true. Using the RepRisk data, we see that high violating countries (e.g. Indonesia) do have significantly higher detected violations than low violating countries (e.g. Peru), therefore the real-world detection bias is not large enough to create an L-shape relationship. Thus, variation in violation detection rates are not a feasible mechanism for generating the L-shape we observe in the coffee and chocolate industries.

E. Certification Effects Mediated by Geographical Sourcing Results

We examine the role played by geographical sourcing in mediating the effect of certification on ESG performance. Certifying firms can take two routes to reduce their violations: 1) source from countries with existing low violations, or 2) build new responsible sourcing infrastructure in challenging regions. The latter is substantially more difficult but is shown to have a positive regional impact in increasing wages, education and living standards (Lee and Tang 2018). However, firms that source from highly challenging regions may have a lower observed reduction in ESG violations due to the difficult nature of the sourcing conditions. Therefore, it is important to control for country-sourcing strategies to remove it as a mediating factor and isolate the benefits of firms with certification. The results below show the comparison of violation rates for company segments with more than 3% certified sales against companies with self-made claims and firms with no responsible sourcing strategies. Table 9 shows the results without controlling for country sourcing and Table 10 shows the results after removing geographical sourcing as a mediating factor. The results show that firms with more than 3% certified sales have an even larger advantage in reducing ESG violations after controlling for country sourcing strategies. This highlights that not accounting for country sourcing would lead to an underestimation of the benefits of certification.

Variable	Estimate	Std. Error
3%+ Certified	−4.66	3.2
Self-made claims	−1.1	2.6
Log(Product Sales)	1.3***	0.3
Year 2	4.8	5.1
Year 3	2.5	5.1
Year 4	3.9	5.0
Year 5	3.3	5.0
Year 6	3.8	4.9
Year 7	9.0	4.9
Year 8	10.2*	4.9
Year 9	10.5*	4.8
Year 10	7.9	5.0
Constant	−14.5**	5.0

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 9 Regression results for yearly ESG violation outcomes for coffee company segments without controlling for country sourcing mechanisms for 398 firm-years.

Variable	Estimate	Std. Error
3%+ Certified	−10.9***	3.1
Self-made claims	−2.96	2.7
Log (Product Sales)	0.9**	0.3
Year 2	5.1	4.3
Year 3	2.8	4.3
Year 4	4.4	4.3
Year 5	3.9	4.2
Year 6	4.2	4.2
Year 7	9.7*	4.2
Year 8	10.8**	4.1
Year 9	11.3**	4.1
Year 10	8.8*	4.1
Côte d'Ivoire	1704***	420
China	1.4	6.0
Colombia	3.5	9.5
Ethiopia	180	335
Kenya	−20.4	53.4
Uganda	−1074	559
India	13	9.6
Indonesia	8.4	9.6
Papua New Guinea	60	58
Vietnam	92.3***	8.5
Costa Rica	−3.3	12.5
Guatemala	−71.6*	36.5
Honduras	7.7	18.8
Mexico	−8.3	6.2
Nicaragua	−153	162
Brazil	−2.6	6.0
Peru	−40.5	23
Constant	−9.4	4.7

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 10 Regression results for yearly ESG violation outcomes for coffee company segments with controlling for country sourcing mechanisms for 398 firm-years.