Auditor Leniency and Participation in Voluntary Forest Certification*

Johanna Isman**

2024-03-13

Job Market Paper

Click here for latest version.

Abstract

Millions of firms seek voluntary certification to signal unobserved quality. This paper investigates how stricter enforcement of certification rules affects firms' participation and quality. I build an empirical model of voluntary certification in which firms choose between competing certifiers. Those certifiers audit quality for the same label with varying levels of rigor. Label owners enforce their rules by excluding excessively lenient certifiers. I estimate the model with novel web-scraped and survey data on the Forest Stewardship Council's (FSC) standard for sustainable wood production. I find considerable differences in certifiers' levels of rigor and suggestive evidence that forest managers are willing to pay substantially more for relatively lenient certifiers. Counterfactuals show that increasing certifiers' minimum level of rigor raises compliance with the certification rules and, thus, quality in certified forests. However, it reduces participation. That trade-off implies ambiguous effects on aggregate quality in certified and uncertified forests. Extreme increases in the minimum level of rigor might have adverse effects, highlighting a general limitation of voluntary certification. For moderate increases, the predicted effects are positive, suggesting that FSC can incentivize more preservation of biodiversity and the carbon stock than it has done so far.

Keywords: Certification, Audit Quality, Voluntary Approaches, Forest Conservation

JEL classification: L15, M42, Q23

*I am particularly thankful to Emmanuelle Auriol, Mathias Reynaert, Stéphane Straub, Pierre Dubois, Isis Durrmeyer, and Eduardo Souza-Rodrigues for their advice and encouragement. I am grateful to numerous further faculty members in Toulouse and Toronto, seminar and conference participants, and Jens Abildtrup, Wesley Blundell, Raymond Fisman, Moritz Hengel, Katrina Mullan, and Benjamin Ouvrard. I thank numerous friends for their help in translating the survey, as well as survey respondents and interview partners from various institutions for sharing industry insights. This work has benefitted from a State grant managed by the National Research Agency under the "Investissements d'Avenir" program (ANR-18-EURE-0021), Emmanuelle Auriol's research grant at the Institut Universitaire de France, and data collection assistance by Jia Wang, research assistant at Assurance Services International. All errors are mine.

** Toulouse School of Economics, 1 Esplanade de l'Université, 31080 Toulouse, France, email: johanna.isman@tse-fr.eu

1 Introduction

Millions of firms seek third-party certification of voluntary standards to signal unobserved quality of their products or services (Dranove and Jin, 2010).[1] Examples range from credit risk and seller quality to organic agriculture and sustainable forestry. Certification mitigates information asymmetries, particularly in the case of *credence attributes*, which consumers and investors cannot observe at a reasonable cost, not even after purchase (Darby and Karni, 1973). Furthermore, certification of sustainability standards can help reduce negative production externalities (Bizzotto and Harstad, 2020), such as the loss of biodiversity and carbon stocks. Yet, well-known non-profit institutions criticize lax enforcement, and several empirical studies find small or no significant differences in the sustainability of certified and uncertified production.[2]

How would stricter enforcement in voluntary certification affect quality in the industry and welfare? Understanding and predicting those effects is not only crucial for private and public institutions that own certification labels.[3] It also helps policy-makers understand the potential and limitations of voluntary certification in regulating externalities.[4] From a theoretical perspective, the impact of stricter enforcement is ambiguous. While it could improve sustainability or other qualities of certified production, it may also reduce participation in certification and, thus, the quality of dropouts (Hui et al., 2023).

This paper studies that empirical trade-off by building and estimating a structural model of voluntary certification in which firms choose between competing certifiers. The model captures the way voluntary certification of most credence qualities is organized. Various certifiers audit forests' quality according to the same standard but potentially with varying levels of rigor. These companies provide the same label in the eyes of consumers and investors. Strategically acting, profit-maximizing firms may, thus, shop for more lenient certifiers (Bar and Zheng, 2019). To counterbalance, label owners appoint an accreditation body that

[1]Quality management system certification according to the ISO 9001 standard alone covers more than a million companies (ISO, 2023).

[2]Examples of such empirical studies are Blackman et al. (2018), Villalobos et al. (2018), Rico-Straffon et al. (2023). See Blackman and Rivera (2011) and DeFries et al. (2017) for reviews of various sectors. Greenpeace International (2021) and the International Consortium of Investigative Journalists (Alecci, 2023) denounce a lack of rigor in certifying land-based commodities.

[3]Voluntary standards set by public authorities include the EU rules for organic certification and the German "Green Button" for sustainable clothing production (European Commission, 2018; Federal Ministry for Economic Cooperation and Development, 2023).

[4]For example, EU institutions discussed whether to recognize certification as proof of compliance when drafting the EU deforestation regulation (European Union, 2023).

inspects and licenses certifiers. To increase the minimum level of rigor, the accreditation body suspends the market access of excessively lenient certifiers, which has side effects on firms' participation and welfare.

I quantify certifiers' rigor differences, firms' preferences for leniency, and effects of counterfactual certifier suspensions with novel data on Forest Stewardship Council (FSC) certification. FSC owns one of the most widespread sustainability standards, covering roughly one-tenth of the global wood production (FAO, 2018). The FSC standard is the most demanding and transparent standard for sustainable forest management.[5] Nevertheless, some FSC certifiers have been accused of leniency, even regarding cases of illegal logging.[6]

This paper contributes to the literature on certification and enforcement in three ways. First, it provides new insights about forest conservation by assessing reasons for the limited effectiveness of certification and the extend to which FSC can improve in view of certifiers' rigor and firms' preferences. Second, the paper extends the broader literature on certifiers' incentives and oversight. While, for example, Duflo et al. (2013a) highlight auditors' incentives problems and the positive effects of stricter control for mandatory standards, I quantify the more ambiguous effects in voluntary certification. Finally, the present paper complements Hui et al. (2023), who study tightened thresholds in the certification of experience quality, which consumers observe and review after purchase. I focus on credence quality, whose certification requires field audits, and I account for the widespread separation of label owners and certifiers in that setting.

To address the research question, I collected novel panel data on FSC-certified forests. I scraped data on all forests that were FSC-certified for at least one year between 2015 and 2019, henceforth referred to as *once-certified* forests. Most importantly, the panel includes yearly observations of participation, the chosen certifier, and the number of violations reported by the certifier during the annual audit. These reported violations of the FSC standard matter since forest firms must correct them to become or remain certified. I extracted these variables from the audit reports published on FSC's website. For the analysis, I focus on various world regions that jointly represent one third of all FSC certificates and provide over 6,000 observations. These regions include the Americas and large parts of Asia. I also conducted an online survey in 24 languages asking forest managers about the certification

[5] Jurists, forest scientists, and NGOs share this view (Clark and Kozar, 2011; Greenpeace International, 2021; Gutierrez Garzon et al., 2020; Ludwig et al., 2014).

[6] See Conniff (2018); Earthsight (2020). FSC consequently asked their accreditation body to "perform a check of the performance of the certification body" (FSC, 2021a), which suggests that they recognize the role certifiers' ("certification bodies"') rigor plays.

fees they pay. The responses provide over 380 certification fee quotes from the regions the paper focuses on, joined with information about the corresponding certificates. I use those data to predict certification fees for all certifiers and forests.

Next, I model voluntary certification as a 3-stage game, played each year. Each world region and year jointly define a market. In the first stage, certifiers set markups as market-level percentages of their marginal cost. In the second stage, forest firms choose whether to become FSC-certified and, if yes, their most preferred certifier. In the third stage, the chosen certifiers audit the forests and detect and report a fraction of forest firms' actual violations of the FSC standard. Correcting those violations is a cost to forest firms' surplus. The reported fraction of violations is a measure of certifiers' audit rigor. Forest firms anticipate certifiers' rigor when choosing a certifier. Conditional on a rich set of audit and forest characteristics, I model both certifiers' level of rigor and the number of firms' actual violations as exogenous variables rather than strategic choices. Certifiers' market-level rigor *differences* depend on their constant rigor types, the distances to their headquarters and inspections by the accreditation body. While the types account for adverse selection, the variation with distances and inspections account for potential moral hazard or behavioral effects in a parametric way.[7]

I estimate the model by backward induction. The abovementioned modeling choices allow me to identify certifiers' rigor differences, i.e., their *relative rigor* compared to a baseline. That is necessary since I observe neither certifiers' level of rigor nor forest firms' actual violations of the FSC standard directly. To obtain reliable estimates of relative rigor from violation reports, I need to account for the potential selection of more or less compliant forest plots into more or less rigorous certifiers. A rich set of audit and forest-level predictors, selected by LASSO regularization, does that to a substantial extent. In addition, I build a control function based on a distributional assumption about the unobserved factors affecting selection and compliance. I combine the approach of Lee (1983) for selection among multiple alternatives in a linear model with the approach of Terza (1998) for binary selection in nonlinear models. To my knowledge, I am the first to do so. The model's estimates suggest that certifiers differ significantly in terms of their relative rigor levels. All else being equal, the least rigorous certifier reports only around one third of the violations that the most rigorous certifier reports. An inspection by the accreditation body is associated with doubling the number of violation reports by increasing the inspected certifier's relative rigor. Variation

[7] In Appendix C.1.5, I discuss to which extent the model nests the possibility of strategic long-term choices of rigor.

in the likelihood of inspections creates within-certifier variation in the relative rigor levels which forest expect for each certifier and market.

In the second stage, I exploit that cross-market rigor variation to identify firms' preference for less rigorous certifiers. To do so, I use demand estimation techniques from the empirical industrial organization literature (Berry et al., 1995; McFadden, 1977). I account for choice variation along various certifier characteristics and unobservable popularity captured by certifier fixed effects. Nevertheless, both prices and expected relative rigor may correlate with unobserved shocks to firms' demand for certifiers. For rigor, the reason is that the accreditation body reacts to changes in market shares by adapting the frequency of accreditation inspections which affects certifiers' rigor. To account for those sources of endogeneity, I use differentiation instruments that capture the degree of market isolation of certifiers in terms of their characteristics (Gandhi and Houde, 2019). But on average, the estimates suggest that forest firms are willing to pay more than 4,000 USD more for a certifier that reports one standard deviation of violations less than their competitor.[8] That is substantial, constituting approximately 40% of the average certification fee and half of firms' average estimated net benefit from FSC certification. Considerable variation in the timing of forest firms' decisions to join FSC or to drop out, coupled with variation in market structures and certifiers' characteristics, helps identifying those net benefits. Firms' estimated willingness to pay to stay with their previous certifier is of similar magnitude as the willingness to pay for leniency, capturing large switching costs that limit shopping for leniency.

Using the estimated model, I conduct two sets of counterfactual exercises. First, I simulate 10 to 500% increases in the minimum level of expected relative rigor worldwide, henceforth referred to as minimum rigor. That is, I shift the lower rigor levels of any certifier and market up to that new level. Such a change mechanically increases the quality of certification. I measure quality as the expected number of violation reports, since more violation reports lead to the correction of more violations and, thus, more compliance. I solve for forest firms' new choice probabilities and certifiers' new prices, keeping everything else constant. The results suggest that the effect on quality among all once-certified forest plots is positive up to increases of approximately 250%. Overall, the relationship between minimum rigor and quality exhibits a hump-shape.

[8]There are open questions about how to derive standard errors in estimating a model with several stages. I present heteroskedasticity and serial correlation robust standard errors, which suggest the significance of the coefficients on prices and certifiers' relative rigor. However, a preliminary set of bootstrap samples suggests that the statistical uncertainty around the model's estimates might be substantial once accounting for the use of generated regressors.

In the second set of counterfactuals, I start by considering the enforcement of increases in minimum rigor. Since many voluntary standards, including FSC, work with multiple certifiers, label owners and their accreditation bodies cannot directly shift minimum rigor. Instead, they do so by suspending excessively lenient certifiers' accreditation and thus their market access. I simulate the accreditation suspensions of the most lenient, the two most lenient and the three most lenient certifiers, by removing them from the choice sets. The trade-off between the quality of certification and participation remains. The overall effects of all three suspension examples on aggregate quality across certified and uncertified forests and welfare are positive. For example, I predict that the suspension of the most lenient certifier leads to a 2.5% increase in aggregate quality, despite almost 6% reduction in participation. I attempt to approximate the potential social benefits and welfare based on survey results on consumers' stated valuation of FSC certification. Despite the reduction in the availability of certified wood and a cost of almost one million USD for forest firms and certifiers, I predict that the suspension of the most lenient certifier could raise welfare by approximately 23 million USD.

Finally, I compare the effects of suspensions to the effects of equivalent increases in minimum rigor for each market. For each set of suspended certifiers, I implement an alternative scenario where I do not remove that certifier but rather shift their rigor to the next most rigorous certifier's rigor in the same market. The set of available rigor levels in each market is then the same as that in the case of a suspension. I find that such equivalent increases in minimum rigor explain less than half of the changes due to suspending the most lenient certifier. Suspensions lead to larger drops in participation but also larger increases in the quality of certification than equivalent increases in minimum rigor. Switching costs and preferences for other certifier characteristics can explain that. Suspension not only changes the available set of rigor levels but also forces forest firms to transfer to another certifier. Transfers are costly, so that more forest firms drop out instead of transfer, when faced with such a situation. At the same time, in the case of a suspension, more forest firms transfer to a certifier that is more rigorous than to the next most lenient certifier. Overall, whether suspensions have better or worse quality and welfare effects than equivalent rigor shifts depends on the characteristics and market shares of the targeted certifier.

Contribution to the literature

The present paper extends a rich literature that investigates reasons for forest degradation and deforestation (Assunção et al., 2023; Balboni et al., 2021; Burgess et al., 2012) and their reduction through private and public policies (Alix-Garcia and Wolff, 2014; Assunção et al., 2023; Simonet et al., 2019; Sims and Alix-Garcia, 2017; Souza-Rodrigues, 2019). In particular, this work complements studies of the effectiveness of forest management certification in preventing forest degradation or deforestation, which certification can do in two ways. First, it can compensate firms' conservation efforts by involving only those forest plots that were already more sustainably managed than average ones before certification. While Goodman et al. (2019) and Kalonga et al. (2016) document such contexts, other studies find no ex-ante difference or even more tree cover loss in later-certified forests (Blackman et al., 2018; Rico-Straffon et al., 2023). Second, certifiers can make forest management more sustainable, for example by reporting violations of the standard and requiring firms to correct them to maintain their certification. Several papers study such causal effects by comparing changes in ex-ante similar forests. Most of those papers find no significant effects (Blackman et al., 2018; Panlasigui et al., 2018; Rico-Straffon et al., 2023; Villalobos et al., 2018), with some exceptions (Miteva et al., 2015; Tritsch et al., 2020). The present paper investigates two potential reasons for the limited effectiveness of FSC and similar certification schemes in some contexts[9] by analyzing the incentives of certifiers and the impact of the voluntary nature of those certification schemes. This work goes beyond existing studies by predicting to what extent stricter accreditation may improve the current level of forest conservation.

More generally, the present paper addresses the economic incentives of inspectors and certifiers. Much of this stream of literature is theoretical (Auriol and Schilizzi, 2015; Bizzotto and Harstad, 2020; Bolton et al., 2012; Mathis et al., 2009; Stahl and Strausz, 2017). Empirically, firms' shopping for less rigorous auditors has been documented for the certification of mandatory environmental or product safety regulations (Chu et al., 2021; Duflo et al., 2013a,b, 2018; Hubbard, 2002) and credit ratings (Becker and Milbourn, 2011; Jiang et al., 2012). To my knowledge, the only quantitative analysis of firms' preference for leniency in voluntary certification is Bar and Zheng (2019), focusing on food safety certification. I identify differences in leniency and firms' willingness to pay for leniency in *voluntary sustainability certification*. There are many reasons to expect important differences in the degree to which firms in this context shop for leniency compared to the settings studied in other

[9] One caveat is that most of those rigorous, quantitative studies look at tree cover loss, whereas FSC may be more effective in other dimensions that are more difficult to measure at a large scale.

papers.[10] More important than such differences are the normative implications. Where participation is mandatory, counteracting shopping for leniency by punishing lenient auditors or assigning them randomly typically has positive effects (Duflo et al., 2013a). In voluntary certification, the outcome is less clear, as firms may stop participating when auditors become too rigorous. I complement Bar and Zheng (2019) by investigating this trade-off. I estimate a structural model that allows me to simulate the quality and welfare effects of suspending lenient certifiers.

Finally, the present paper joins Hui et al. (2023) in empirically investigating the trade-off between stringency and participation in voluntary certification. Hui et al. (2023) exploit rich data and a change in eBay's seller certification to analyze the effects of increased stringency on the distribution of quality and other market outcomes. My data and context do not allow me to estimate such a rich model incorporating moral hazard. However, I complement that paper by focusing on a certification standard that is very different from eBay's seller certification but representative of many voluntary schemes, particularly those addressing ethical and environmental issues. Certification on internet platforms such as eBay typically reveals experience attributes, observed after purchase. Platforms can certify easily by summarizing the reviews of previous customers. I investigate the trade-off between stringency and participation in the voluntary certification of credence attributes, which is not even observable after purchase. The certification of such quality attributes often requires extensive field audits, which has led label owners to outsource certification to multiple certifiers. As my paper shows, this structure has significant consequences for the enforcement of increased stringency and consequent effects on quality. Moreover, I analyze the effects with a view towards externalities, which matters for interpreting welfare effects.

Overview

The remainder of this paper is structured as follows. Section 2 describes the institutional setting. Section 3 presents the data. Section 4 outlines the model of violation reporting,

[10]Ex-ante, it is not clear if shopping for leniency is exacerbated or mitigated in the context of voluntary certification. On the one hand, punishment for leniency can be more severe in mandatory certification. Credit rating agencies also face more considerable risks of losing their reputation as (i) their brand is more observable to clients than are the brands of certifiers working for the same label as FSC, and (ii) clients are likely to more appropriately judge differences in their overall accuracy ex-post. On the other hand, intrinsic motivation may be more present in voluntary certification than in mandatory certification and credit ratings. Food safety differs from environmental protection in that the former is more of an experience than a credence attribute, similar to credit ratings.

demand for certifiers, and pricing. Section 5 details the estimation and presents and discusses the results. Section 6 covers the counterfactual analysis. Section 7 concludes the paper.

2 Institutional setting

This section describes the key features of the FSC certification system. Many of these features are typical for most voluntary certification schemes of credence qualities and motivate the model's focus. Other features are more specific and will help interpret the quantitative results of the estimated model.

2.1 Forest management units

Approximately 10% of the global forest area is certified according to sustainable management standards (UNECE/FAO, 2019). This paper analyzes certification decisions over time at the forest management unit (FMU) level. FMUs are equivalent to establishments in other industries. They are defined forest areas managed by the same firm according to a joint plan (FSC, 2017).[11] A certificate can cover an individual FMU or a group of FMUs, jointly responsible for compliance. I refer to the whole entity covered by one certificate as an FMU but account for jointly certified groups of FMUs in the analysis.

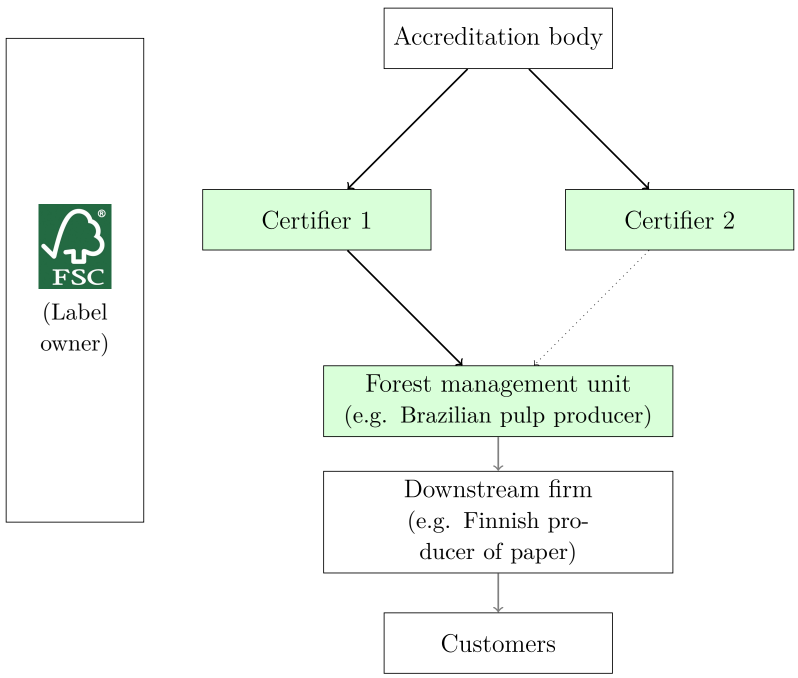
FMUs seeking certification can choose between two main, globally recognized sets of standards: the Forest Stewardship Council (FSC) standard and a wide range of other standards recognized by the Programme for the Endorsement of Forest Certification (PEFC). This paper focuses on FSC for two reasons. First, only FSC publishes audit results, which are indispensable for empirical analysis. Second, the FSC standard is the most demanding standard, also compared to national regulations.[12] The analysis thus considers all other standards as less demanding outside options. FSC certification covers roughly one tenth of the global wood production and approximately 4% of the world's forest area, equivalent to the size of Mexico.[13]

[11]FMUs are typically predefined administratively for publicly owned forests, while they can be more flexibly adapted by private owners.

[12]Jurists, forest scientists, and NGOs share this view (Clark and Kozar, 2011; Greenpeace International, 2021; Gutierrez Garzon et al., 2020; Ludwig et al., 2014).

[13]See Appendix Figure A1, FAO (2018) and UNECE/FAO (2019).

Figure 1: Illustration of the FSC certification scheme



Notes: Black arrows symbolize monitoring through audits and inspections, the dotted arrow counterfactual monitoring if Certifier 2 would have been chosen. Gray arrows symbolize the supply chain. The blue color highlights the firms which are the focus of this paper. The position of FSC on the left symbolizes the relevance of its standards and label for all levels of the monitoring and supply chain.

2.2 The FSC certification standard

Figure 1 illustrates the FSC certification scheme, which is organized similarly to most other voluntary certification schemes.[14] After choosing the FSC standard, the FMU chooses among certifiers approved by FSC's accreditation body. The chosen certifier decides on the FMU's certification based on audits, as described below. Downstream firms can use the FSC label on wood, paper, and other forest products if they source sufficiently from FSC-certified forests.[15] This paper restricts attention to FMUs' certification according to FSC's forest management standard, henceforth referred to as *the FSC standard*.[16]

The FSC standard regulates the environmental and social externalities of forest management rather than traditional quality characteristics of forest products. The quality on which the FSC standard focuses, thus, involves credence rather than experience attributes, i.e., not even observed after purchase (Darby and Karni, 1973). The rules include reducing the amount of logging in areas of high conservation value, sustainable logging volumes, and worker safety equipment. All rules stem from the following ten principles (FSC, 2015):

1. Compliance with Laws
2. Workers' Rights and Employment Conditions
3. Indigenous Peoples' Rights
4. Community Relations
5. Benefits from the Forest
6. Environmental Values and Impacts
7. Management Planning
8. Monitoring and Assessment
9. High Conservation Values
10. Implementation of Management Activities

The standard has national and regional versions, but all versions must align with a common set of core rules.[17] The violations considered in this paper are mostly violations of those core rules, but I also control for regional differences.

[14]In particular, major voluntary certification schemes let firms choose their certifier, including those covering EU organic production and ISO standards. Fair Trade is a rare exception.

[15]That certification follows FSC's chain-of-custody standard.

[16]Most certified FMUs are also certified according to the chain-of-custody standard since they sell their forest products. They typically choose their certifier and pay their fees jointly for both standards. Since the audit results for chain-of-custody certification are not public, I can only account for the impact of the forest management standard on a firm's certifier choice. I exclude a small number of FMUs that are not simultaneously certified according to the chain-of-custody standard.

[17]These core rules include the "Principles and Criteria" and additional policies and procedures specifying

Firms' costs and benefits from FSC certification are heterogeneous. The costs include opportunity costs from compliance with the standard and certification fees for the certifiers.[18] There is little public information on such costs. FSC does not regulate certification fees, except requiring that these fees follow a general pricing schedule.[19] Small-scale surveys suggest a total cost of approximately 6 to 40 USD per hectare and year for FMUs of less than 4,000 hectares and 0.07 to 0.5 USD for large FMUs of more than 400,000 hectares in the Americas, with similar costs in Japan (Chen et al., 2010; Cubbage et al., 2009; Sugiura and Oki, 2018). However, the design of these surveys suggests that the responses do not include all opportunity costs from compliance with the FSC standard, such as reduced revenues due to restricted logging rates.[20]

The economic benefits of FSC certification vary across markets and firms. FSC certification often secures continued business with downstream buyers.[21] Some markets provide a direct premium of up to 56% for FSC-certified wood.[22] According to a meta-analysis covering mostly advanced economies, the median consumer is willing to pay a price premium of approximately 10% of the retail price (Cai and Aguilar, 2013). A survey commissioned by FSC suggests that about every second consumer in 33 countries recognizes the FSC label and approximately every fourth consumer is willing to pay more for an FSC-certified product than for an uncertified product (FSC and IPSOS, 2023). Overall, small-scale surveys and case

their interpretation. These rules are written and voted on by an assembly of FSC members from environmental NGOs, the wood industry, and representatives of communities from forest areas.

[18]These costs also include anegligible FSC membership fees, which amount to only 55-160 USD for a median-sized certified forest, depending on the forest type (FSC, 2016a).

[19]The rules state that a general schedule of certification fees should be publicly available (FSC, 2015), but FSC does not enforce this requirement entirely. Instead, interested FMUs have to ask for a quote from certifiers. One certifier states that the fees depend on the "operation's size, geographic locations, and the complexity of factors such as forestry activities, high conservation values, stakeholder relations, etc." (NEPCon, 2019). An interviewed association of certified FMUs in Canada suggested that the fees are typically not negotiated.

[20]In Cubbage et al. (2009), only four out of 14 respondents included an estimation of the cost of "changes required to get/maintain certification". Given the cost quoted by those respondents (0.059 USD per hectare at the median) and the formulation of the question, it is unlikely that this cost includes all foregone profits. The situation is similar in Sugiura and Oki (2018). For Sweden, Villalobos et al. (2018) report only average direct payments for FSC and PEFC certifiers, suggesting hectare rates of 1-2 USD for recertification and 10-30 cents for annual surveillance audits.

[21]See Araujo et al. (2009); Overdevest and Rickenbach (2006); Galati et al. (2017); FSC (2018a); Holopainen et al. (2015). Forest management certification is sometimes a condition for participation in public procurement, as is the case in Germany (Ludwig et al., 2014).

[22]Frey et al. (2018) find increased revenues of 10-30% in Vietnam and Kollert and Lagan (2007) report a certification premium of up to 56% for high quality tropical wood in Malaysia. FSC certification may also be a precondition or facilitator for public subsidies (Visseren-Hamakers and Pattberg, 2013). However, most Japanese firms surveyed by Sugiura and Oki (2018) and a Canadian smallholder I interviewed do not see any direct economic benefits. FSC is by far not the only label, for which not all certified producers receive a price premium (Dragusanu et al., 2014; Subervie and Vagneron, 2013).

studies suggest that additional revenues exceed the total cost of certification for only roughly every second certified FMU.[23] For other FMUs, less tangible benefits drive participation. For example, certification makes it easier to obtain licenses to import wood products into countries with timber legality regulations, such as EU member states.[24] Surveys suggest that many firms seek certification to improve their reputation, potentially hoping for more tangible benefits in the future. Social prestige concerns and intrinsic benefits due to learning about sustainable practices also play a role.[25] For state-owned firms, certification is often a political decision.[26]

2.3 FSC certifiers

FSC certifiers are both for-profit and nonprofit companies. They evaluate FMUs' compliance with the FSC standard in annual *surveillance audits* and in more extensive *(re)certification audits* in the first year and every fifth year (FSC, 2015).

Violation reporting: In every audit, certifiers report minor and major violations of compliance with the standard. *Major violations* are significant failures of compliance with FSC's core rules and the focus of this paper.[27] Examples include insufficient monitoring to protect endangered species or a management plan with logging volumes that would not sustain the forest cover. Many violations concern workers' rights, such as insufficient protection clothing, and insurance. Others relate to the local community and the state, such as the omission of consulting indigenous communities before making the harvest plan.[28] Reporting such violations is the main way certifiers enforce compliance with the FSC standard. If certifiers find five or more major violations in one audit, they must suspend the FMU's certificate (FSC,

[23]See Cubbage et al. (2009); Frey et al. (2018); Kitchoukov et al. (2019); Sugiura and Oki (2018); Owari and Sawanobori (2007).

[24]See Holopainen et al. (2015), Gavrilut et al. (2015) and the recognition of an FSC certification body as providing sufficient evidence for timber legality in Cameroon (FSC, 2016b).

[25]See the surveys by Araujo et al. (2009); Overdevest and Rickenbach (2006); Galati et al. (2017) and Paluš et al. (2021).

[26]This has been highlighted in interviews with industry insiders in Germany and Austria.

[27]If not specified otherwise, the violations in this paper are always major violations. Note further that certifiers must upgrade minor violations to major violations if FMUs do not correct these violations within a year.

[28]These examples come from the descriptions of 35 violations publicly reported in 20 audits randomly drawn from all observations with at least one violation in the forest unit panel (described in Section 3.1). In a sample of 110 audit summaries from Brazil in 2016, Rafael et al. (2018) find that most of the violations relate to environmental issues, community relations, and workers' rights. Furthermore, Blackman et al. (2017) examine the audit summaries of 35 FMUs in 2000-2013 in Mexico and find that most violations relate to social and economic-legal issues rather than environmental ones.

2009b). If they find under five, they must check their correction in an additional audit after three months and suspend the certificate otherwise.[29] Most FMUs, therefore, correct their reported violations. For a given level of compliance ex ante, more violation reports, thus, translate into greater compliance and, hence, fewer negative externalities and higher quality among the certified.

For FMUs, violation reports generate the cost of an additional audit plus the opportunity cost of correcting the violation or losing the certificate. There is no estimate of those costs in the literature thus far. The additional audit needs only to check that the reported violations have been corrected. Hence, the fee is likely a minor fraction of the usual annual certification fee. Correcting violations requires that the violation be ended and that procedures are adjusted to prevent similar violations in the future. The cost of such corrections varies. Among the violation reports from 20 randomly drawn audits, most corrections require relatively low levels of investment, such as purchasing a missing protection cloth for loggers. About one-third of the corrections seem to require medium-range investments, such as assessing a road's environmental impact before construction. In addition to the direct cost, roughly half of all violation reports seem to bear a small risk of much higher opportunity cost in the future, for example, if an environmental impact assessment concludes that the FMU cannot build a road due to the presence of protected species. A minimal share of violation reports requires the limitation of harvests and may, thus, lead to opportunity costs of over a hundred thousand USD.[30]

Certifiers have some discretion in auditing and reporting violations. For example, they have only to check a certain number of criteria and have some freedom in choosing those criteria. FMUs can transfer to another certifier any year, even though doing so is not encouraged by the FSC.[31] They can only transfer after correcting any major violations reported by the current certifier (FSC, 2010). FMUs obtain guidance and information from other companies when choosing their certifiers (Sugiura and Oki, 2018). Industry insiders suggest that the relevant factors for FMUs' certifier choice are fee differences, existing relationships with certifiers, efficiency, local presence and expertise of the certifier, but also differences in rigor.

[29]Certifiers may lift the suspension if the FMU corrects the major violations within a year. Otherwise, the certificate is withdrawn (FSC, 2015). An FMU can apply for a new certificate after the withdrawal of a previous certificate (FSC, 2021c), but doing so requires the correction of all major violations.

[30]From a case study by the NGO Earthsight (2020), it can also be crudely estimated that an FSC-certified Ukrainian forest enterprise was able to gain around 100,000 USD when illegal logging practices were undetected by their FSC certifier.

[31]If FMUs transfer to another certifier more than once in the five years between (re)certification audits, then they have to perform an additional (re)certification audit, implying increased costs.

Accreditation of FSC certifiers: FSC certifiers must be licensed by FSC's accreditation body, Assurance Services International (ASI).[32] ASI assesses certifiers' competence and compliance by reviewing documents and inspecting their audits of FMUs and their offices at least once per year per certifier (ASI, 2019). For inspections of audits, ASI either is present as an observer while the certifier audits or inspects the FMU afterward to compare the results with those in the certifier's report. I refer to both cases as *accreditation inspections*. If ASI finds that certifiers do not fulfill their duties, it may suspend their accreditation globally or for some regions. In such cases, the affected FMUs have six months to contract new certifiers to keep their certification valid (FSC, 2015, Art. 1.1.3). Despite certifiers' accreditation, activists and journalists have accused certifiers of issuing certificates despite severe violations of the FSC standard (Alecci, 2023; Earthsight, 2021; FSC Watch, 2020). In addition, FSC member organizations demanded investigating threats to the impartiality of certifiers to maintain the FSC's credibility (FSC, 2019).

3 Data and summary statistics

I collect information from various sources to construct three novel datasets covering 2015-2019.[33]

1. Forest management units' (FMUs') characteristics, yearly demand for certifiers, and audit results.
2. Certifier characteristics at the market-level.
3. Certification fees from an anonymous survey of certified FMUs.

The FMU and the certifier-market data jointly form the main dataset, henceforth referred to as the FMU-certifier panel. The final datasets cover the following regions, as shown in Figure 2: the Americas except for Venezuela; Southern and Eastern Africa; Asia except for China and Malaysia; Oceania; and Eastern and Southern Europe except for Russia, Ukraine, and Italy. I exclude observations from other regions due to insufficient representation in the price panel or insufficient variation in market shares to identify the model parameters.[34] The

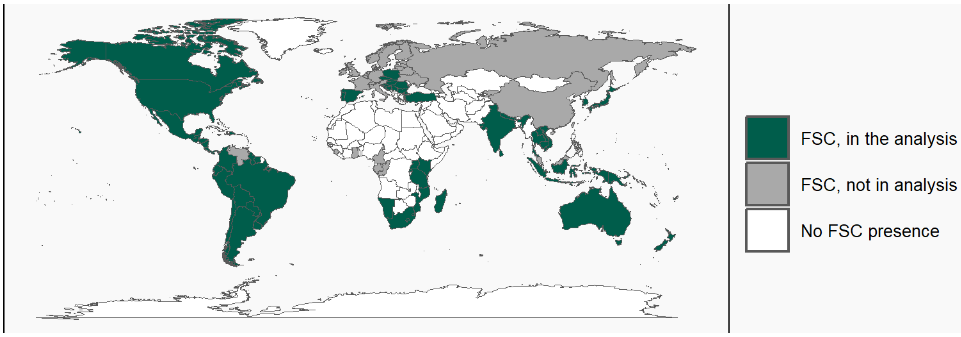
[32]ASI is a company that conducts certifier assessments for sustainability standards in various industries (ASI, 2022). FSC is the unique shareholder of ASI but delegated "full business control" to ASI's supervisory board in 2017.

[33]I exclude the years before 2015 since the number of missing audit documents is high in those years. I exclude 2020 to avoid any disruptions caused by the COVID-19 crisis.

[34]In Western Europe, Sweden and Finland, Western and Middle Africa (particularly the Congo Basin), and the countries of the Commonwealth of Independent States (CIS, particularly Russia and Ukraine), there

included regions represent one-third of all FSC certificates. Audits in these regions result in more average numbers of violation reports than those in the excluded regions (see Appendix Figures A2).

Figure 2: FSC presence and the regional focus of this paper



The datasets include all FMUs certified for at least one year in 2015-2019.[35] Throughout the paper, I refer to such FMUs *once-certified* FMUs. There is no comprehensive dataset for those FMUs that have never been certified. The population this paper considers is, thus, a selected sample from the population of all FMUs. The results do not necessarily extend to the entire population of FMUs. I do not observe all relevant variables for all FMUs throughout 2015-2019, but I impute missing observations, as outlined in the following sections.

3.1 FMU panel

I construct a panel of yearly observations of once-certified FMUs primarily from information on the FSC website (FSC, 2020). Each FMU-year observation corresponds to one *choice situation*, that is, one choice of whether to participate in FSC certification and, if yes, the choice of one certifier. FSC publishes information about all certificates issued in the past, including the first issue date, termination date, country, size of the certified area, group certificate members, products and tree species, and audit documents (see Appendix Figures A3 and A4). I web-scraped that information and downloaded all audit documents in May

is too little variation in firms' certifier choices to precisely identify market-level alternative constants. In Italy, no firms chose the outside option, i.e., ended their FSC certification or were not certified yet, in 2019. CIS countries, the rest of Northern Europe (particularly UK), and China are not well represented in the survey providing the price panel. In Malaysia and Venezuela, important country-level statistics are missing, but these countries have very few certified FMUs.

[35] I include those FMUs whose certificates became valid only in 2020 since they may still have received their first audit and made their certifier choice in 2019.

and June 2020. The information does not specify the owner of the FMU.[36]

From the audit documents, I extract the certifier, year, type of audit and the number of major violation reports. To do so, I write an algorithm using regular expressions. One-third of the numbers of violation reports are also checked manually, to ensure accuracy given the heterogeneity of document formats and languages. The extracted data cover 80% of all choice situations in which FMUs were certified, and include choice situations of 93% of all once-certified FMUs. I impute firms' certifier choices for the missing choice situations following reasonable assumptions. For example, suppose I have no observation for a year preceded and followed by a year with the same certifier. In that case, I assume that the FMU chose that certifier in the missing year since FMUs never switch the certifier two years in a row. Excluding imputed choices dramatically drives up the number of "unstructural zeroes", which can lead to more substantial bias (Briesch et al., 2008). Appendix Sections B.1.1 and B.1.2 detail the reasons for missing data and the data construction process, including the favorable results of accuracy tests.

I match the panel with additional data from various sources. First, I add indicators of corresponding accreditation inspections from the accreditation body's website (ASI, 2022). Second, I add a rich set of annual characteristics of the FMU's country to account for differences in the cost of compliance and demand for certification, including wood product trade values, a corruption index, and the number of certified downstream firms. See Appendix B.1.4 for details. Third, I include an indicator for FMUs which are plantations. For about every fourth FMU, I find such information in audit documents or a recent FSC register. I use these FMUs to train a logit model to classify the remaining FMUs. I select relevant features from the available FMU characteristics with repeated cross-validation and LASSO regularization.[37]

Table 1 summarizes the most important characteristics of FMUs, their choices and audits. The data feature time variation in FMUs' participation in FSC certification which helps identify FMUs' willingness to pay for the certification: in 13% of the choice situations,

[36]This missing information implies that I cannot account for the joint ownership of various FMUs. That is, I have to assume independence of unobserved shocks from the benefits of certification for different FMUs with the same owner. This assumption probably does not affect the analysis greatly since separate certification suggests separate management, as owners have an interest in certifying jointly managed units through one certificate, given that joint certification lowers costs. Only 3% of the certified entities have names which appear on multiple certificates. The maximum number of certificates with the same name as the certified entity is 6. However, entities with different names may still have a joined owner.

[37]The prediction accuracy in a test sample is about 68%. I also attempt K-Nearest Neighbor matching, but the classification error is much smaller with the logit model.

Table 1: Summary statistics of the forest unit panel

Statistic	N	Median	Mean	St. Dev.	Min	Max
Not yet certified	6,250	0	0.13	0.34	0	1
Termination of last year's certificate	6,250	0	0.11	0.31	0	1
Transfer to another FSC-certifier	4,596	0	0.03	0.17	0	1
No. of violation reports (by year)	3,810	0	0.60	1.65	0	33
At least one violation report	3,810	0	0.25	0.43	0	1
Five or more violation reports	3,810	0	0.02	0.14	0	1
Audit inspected by accreditation body	3,810	0	0.03	0.18	0	1
Yrs. with FSC cert.	6,250	5	5.98	5.48	0	26
(Re)certification audit	6,250	0	0.35	0.48	0	1
Classified as plantation	6,250	0	0.24	0.42	0	1
Logging is FMU's primary activity	6,250	1	0.91	0.29	0	1
Certified area in 1000 ha	6,250	12.98	110.09	342.75	0.005	5,986.68
Group certificate (vs. individual)	6,250	0	0.24	0.43	0	1
No. of certificate members	6,250	1	6.64	54.78	1	1,563

FMUs do not choose FSC certification, yet, and 11% are terminated certificates. While most FMUs stick to their initial certifiers, 3% of all choice situations involve transfers to other certifiers.

Since there is no reliable observed measure of differences in certifiers' audit rigor,[38] this paper exploits variation in major violation reports to identify such differences. The numbers of violation reports per year range from 0 to 33. The median is zero. Appendix-Figure A5 shows that the distribution resembles a Poisson, with a longer tail of extreme observations. The average numbersof violation reports varies across certifiers, as Appendix-Figure A6 demonstrates. If certifiers audited the same population of FMUs, those averages could be used as a proxy of their rigor. However, this is not the case, as the examples in Appendix-Figures A7 and A8 illustrate. If less compliant FMUs select into more lenient certifiers, they might end up reporting more violations than more rigorous certifiers. This problem motivates the construction of a model that disentangles these differences. The model will exploit variation in the assignment of inspections by the accreditation body. Such inspections appear in 3% of all audits. They are associated with increased numbers of violation reports, as Appendix-Figure A20 (a) suggests.

The model will account for FMUs' characteristics which might affect their choices and violations. Table 1 summarizes a few important examples. Most once-certified FMUs are natural

[38]The only observed signals of rigor differences are past suspensions, complaints by NGOs, and mistakes that FSC's accreditation body reported from inspections in 2017-2019. However, even the accreditation body does not see those incidents as reliable measures of rigor, due to the small number of inspections, among other things.

forests used for logging. The majority cover large areas of tens of thousands of hectares. A long tail goes up to more than 5 million hectares. Approximately one-fifth are group certificates. The median group has 6 members. The oldest certificate has been valid for 26 years, with the median being four. A more extensive (re)certification audit is needed in about one-third of choice situations.

3.2 Market definition, choice sets and certifier-market panel

I define a market as the combination of the year and a world region. Every year, there is a new market since certifiers tend to charge certification fees annually and FMUs can switch certifiers annually. A market covers a world region defined by two criteria. First, a region corresponds to a United Nations subregion (United Nations, 2023). I exclude certifiers from a region's choice set if they did not certify any FMU in the region in 2015-2019. If they did not certify in the region in the first or last one to four years, I exclude them from the region's choice set for those years, following Briesch et al. (2008). I do so since certifiers are not available in all countries for which they are accredited. In contrast, an office presence in one country can cover countries nearby without public information concerning their availability. Second, if a certifier did not have accreditation for a subset of countries in a given region, I split the region accordingly to have a common choice set within each market. The data cover ten world regions over five years and, hence, 50 markets.

Table 2: FSC certifiers active in 2015-2019

Certifier	Global entry	# active markets	Share across markets 2015 (%)	Share across markets 2019 (%)	Head-quarters	# countries with offices	Type	Company revenue 2020 (MM USD)	Past suspen-sions
1	2011	50	38	29	DNK	60	Non-profit	23	0
2	2005	26	6	4	FRA	75	Traded	5228	2
3	2005	29	3	5	NLD	64	Private	17	0
4	2000	24	5	8	DEU	9	Private	159	0
5	1995	40	14	14	GBR	50	Non-profit	30	0
6	1995	40	15	16	USA	10	Private	38	0
7	1995	45	15	16	ZAF	123	Traded	5962	1
Small certifiers	1998 (earliest)	12	3	8	DEU (largest)	137	Mixed		0

I collect information about certifiers from their websites and the accreditation body. Table 2 presents their main characteristics. Throughout the paper, I focus on the seven largest certifiers who certified over 90 percent of once-certified FMUs.[39] I merge the remaining

[39] I further consider two, originally separate certifiers, Certifiers 1(a) and 1(b), as a single certifier since

certifiers into a competitive fringe, as their certification activities are too small to estimate separate parameters precisely. Most large certifiers have been active since the early years of FSC, but the certifier with the highest market share did not join until 2011. Certifiers differ in terms of their market presence. Market shares vary only slightly over time but greatly across regions, as Appendix Figures A9 show. The consequent variation in FMUs' choice sets is helpful for the model's identification. Most certifiers have headquarters in the Global North but offices in many countries. Two large certifiers are nonprofits specializing in sustainability standards, mainly agriculture and forest management. Most for-profit certifiers do business in a broader range of industries and have higher revenues. The two publicly traded certifiers are the only certifiers that the accreditation body suspended in 2010-2020. Certifier 7's suspension in 2011 led to its permanent exclusion from Brazil. Certifier 2 was suspended worldwide for a few months in 2016-2017 and has been excluded from Russia since 2015 (ASI, 2020).

3.3 Price panel

As FSC certifiers do not publish their certification fees, I surveyed FMUs.[40] I e-mailed the survey to all FMUs with valid certificates in June 2020 for whom the audit reports or their websites included an e-mail address. Owing to a response rate of 21%, I obtained 387 fee quotes and corresponding FMU characteristics from the world regions on which this paper focuses. I convert the quotes into real USD prices using the World Bank (2020)'s currency conversion factor and the US paper and wood pulp producer price index (FRED, 2020), with 2015 as the base year. I conducted the survey anonymously and asked about FMUs' characteristics only in broad categories to encourage participation. In particular, I asked about the size of the certified FMU in six categories from "< 1,000 ha" to "> 500,000 ha" and for the 5-year-interval in which the FMU was initially certified. In the analysis, I replace those categorical responses with the within-category averages by market from the FMU panel, i.e., the whole population of once-certified FMUs, as a numeric characteristic. Appendix B.3 describes the design and outcomes in detail.

Table 3 presents summary statistics of the fee quotes and respondents' most important characteristics. There is an enormous amount of variation in certification fees. The median

Certifier 1(b)'s certification department was acquired by Certifier 1(a) in 2018. Almost all staff working on FSC certification at Certifier 1(b) moved, and certificates transferred quite easily, as industry insiders suggest.

[40] Certifiers were unwilling to share their fees upon request.

Table 3: Summary statistics of the price panel

Statistic in 1000 USD	N	Median	Mean	St. Dev.	Min	Max
Annual cert. fee in 1000 USD, PPI adj.	387	7.37	9.41	8.24	0.67	58.20
Annual cert. fee in USD/ha	387	0.39	1.48	2.79	0.01	20.05
Certified forest area in 1000ha (market-mean by category)	387	22.84	57.66	155.30	0.35	1,399.48
Classified as plantation	383	1	0.56	0.50	0	1
Group certificate (vs. individual)	387	0	0.35	0.48	0	1
(Re)certification audit	384	0	0.28	0.45	0	1
Yrs. with FSC cert. (market-mean by category)	382	8.88	8.92	6.18	0.00	23.00

fee among survey respondents is approximately 7000 USD, roughly 0.4 USD per hectare. Fees range from a few cents to more than 25 USD per hectare. Such fees are consistent with case studies (Chen et al., 2010; Cubbage et al., 2009; Sugiura and Oki, 2018). Respondents represent the variety of FMU types and regions quite well, as detailed in Appendix B.3.2. They tend to be slightly larger than in the whole population, are more often plantations, and have been certified for a longer time. The model estimation will account for such sample selection, but the selection is most likely helpful for the response precision, as managers of larger FMUs probably record fees paid for certification even more reliably than managers of smaller FMUs. I check for consistency in the responses in various dimensions and exclude a minimal number of unreliable responses. Many respondents raised additional confidence in their motivation to respond thoroughly through additional, detailed comments they made in writing.

I use the price panel to predict prices for all certifiers and FMUs. Specifically, I predict prices as the product of a certifier-market-level price and an FMU-market-level price factor. I motivate that in the next section, when describing the model. I select regressors of total prices from a large set of variables which vary either by FMU i and market t ($\boldsymbol{f_{it}}$) or by certifier j and market t ($\boldsymbol{x_{jt}}$). The certifier-variant variables include functions of certifiers' market shares to capture variation in markups. I choose relevant regressors using repeated cross-validation with LASSO regularization (Hastie et al., 2001). I regress log prices per log mean certified forest area by category and market, $\log(\overline{area}_{it})$, on the selected regressors. I then use the estimated coefficients and the FMU-certifier-panel to predict certifier-market-level prices p_{jt}, the individual price factors c_{it} and total prices p_{ijt}.

$$\log\left(\frac{p_{ijt}}{\log(\overline{area}_{it})}\right) = \boldsymbol{f'_{it}}\tilde{\boldsymbol{\rho}}^{\boldsymbol{f}} + \boldsymbol{x'_{jt}}\tilde{\boldsymbol{\rho}}^{\boldsymbol{x}} + \tilde{\epsilon}^{p}_{ijt} \tag{1}$$

$$\mathbb{E}[p_{jt}|\boldsymbol{x_{jt}}] = \exp(\boldsymbol{x'_{jt}}\tilde{\boldsymbol{\rho}}^{\boldsymbol{x}}) \qquad \mathbb{E}[c_{it}|\boldsymbol{f_{it}}] = \log(\overline{area}_{it})\exp(\boldsymbol{f'_{it}}\tilde{\boldsymbol{\rho}}^{\boldsymbol{f}}) \qquad \mathbb{E}[p_{ijt}|\boldsymbol{f_{it}},\boldsymbol{x_{jt}}] = c_{it}p_{jt} \tag{2}$$

However, the price panel is not a representative but a selected sample of the prices offered to all FMUs by all available certifiers. I only observe prices of (1) FMUs that decided to participate in the survey and (2) for the certifiers chosen by the FMUs. The included regressors already account for many of the characteristics that influence those choices and prices. However, remaining sample selection due to unobserved factors might still cause bias in the prediction exercise. Fortunately, I can control for that selection bias based on assumptions about the joint distribution of the error terms in the different models. First, I assume that the error terms of the utility from the choice to participate in the survey and the choice of the certifier are independent, conditional on the other price regressors. I can then control for those two sources of bias separately. Second, I assume that the error term of the survey participation and the pricing prediction models are jointly normal. The Inverse Mills Ratio of the utility of survey participation predicted from a probit model then accounts for the bias from that selection (Heckman, 1979), as outlined in Appendix B.3.3.1. The control functions derived by Lee (1983) account for bias due to the choice of the certifier.[41]

Table 4: Summary statistics of the price predictions

Statistic	N	Median	Mean	St. Dev.	Min	Max
Predicted price p_{ijt}	34,794	7.49	9.44	7.68	0.31	128.29
Predicted certifier-market-level price p_{jt}	34,794	3.21	3.33	1.33	1.27	11.67
Predicted FMU-market-level price factor c_{it}	34,794	2.41	2.90	2.12	0.18	25.80

Table 4 shows the price predictions for all FMU-certifier-year observations. While the outlined prediction procedure aims at a good balance between bias and explained variance in out-of-sample prediction, the predictions still explain 72% of the in-sample variation in the price quotes. Differences between FMUs and markets matter most, but certifier-variant predictors still explain 20% of the predicted variation. At all quartiles, the predicted prices are lower than the prices quoted in the survey. This difference is not surprising. FMUs that pay higher prices might see more benefits in FSC certification and contribute more willingly to research. The difference suggests that controlling for selection into the survey is important.

[41]The control functions follow the form in Equation (3.7) in Lee (1983). They are based on the transformation (18) of the nested logit error, as described in Section 5.1.1 of this paper.

4 Empirical model

I build an empirical model of voluntary certification. I apply the model to FSC certification of forest management units (FMUs), but the main ideas are applicable for many certification schemes. The model describes how certifiers' rigor may impact violation reporting; That part of the model allows to identify a proxy of the differences in certifiers' rigor, tackling one of the main challenges of this paper: certifiers' rigor is not observed. Next, the model illustrates the role of certifiers' rigor in FMUs' participation and certifier choices, both directly and indirectly through certifiers' markups. It provides a framework to simulate how an increase in minimum rigor levels through stricter accreditation can affect violation reporting and participation.

There are T markets, defined by year and world region. In each market t, FMUs and certifiers play the following stage game. In Stage 1, each certifier j sets their certification fee. In Stage 2, each FMU i decides whether to participate in FSC certification and chooses a certifier. In Stage 3, certifiers audit FMUs and detect and report violations of the FSC standard.[42] This section describes the stages in reverse order.

4.1 Stage 3: Violation reporting

Violation reports v_{ijt} are an exponential function of certifier, audit and forest management unit (FMU) *types* plus a shock:

$$v_{ijt} = \underbrace{r_j \exp(\boldsymbol{x}_{ijt}^{v\prime}\boldsymbol{\omega})}_{\tilde{r}_{ijt}} \underbrace{\exp(\boldsymbol{f}_{it}^{v\prime}\boldsymbol{\gamma} + \eta_i)}_{v_{i1t}^0} + \epsilon_{ijt}^v \tag{3}$$

The empirical distribution motivates the functional form.

r_j denotes a fixed effect of certifier j. It captures certifier differences in violation reporting that vary across neither regions nor years.[43] It is a proxy for certifiers' *rigor type*. Rigor types might reflect long-run decisions and expertise, for example, due to the legal status and intrinsic motivation of owners and managers. The intercept γ_0, contained in $\boldsymbol{f}_{it}^{v\prime}\boldsymbol{\gamma}$, absorbs the rigor type of one of the certifiers. I consider Certifier 1 as that *baseline certifier* since

[42] If an FMU chooses an FSC certifier in the subsequent year, I conclude that it corrected the reported violations, as is necessary under the FSC standard.

[43] The model applies to a few years, a period in which companies are unlikely to change drastically. In the application of this paper, five years are considered.

Certifier 1 is present in all markets and has the largest global market share. I normalize r_1 to 1. r_j then measures the fraction of violation reports of Certifier 1 reported by certifier j, everything else equal.

Certifiers' overall rigor may vary across audits. It depends on $\boldsymbol{x}^{\boldsymbol{v}}_{\boldsymbol{ijt}}$, a vector of two audit characteristics: an indicator for an accreditation inspection and the distance between the market and the certifier's headquarters. These factors reflect hypotheses shared by industry insiders. First, certifiers tend to report more violations when the accreditation body inspects the audit they conducted. Second, the more distant the market is from the headquarters, the lower the degree of quality control of auditors and, hence, the level of rigor.

Violation reports also vary with certifier-invariant characteristics of the audit, the FMU and its market. The vector $\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}$ includes observable characteristics, while η_i is an unobserved compliance type. $\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}$ can affect both FMUs' actual violations and certifiers' rigor. For example, FMUs that have been certified for many years might violate the standard less often than newly certified FMUs, but certifiers might also detect more violations among these FMUs, as they have more information from previous audits. I assume only that the rate at which $\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}$ affects rigor does not differ systematically across certifiers.

Overall, the number of violation reports is the product of relative rigor, $\tilde{r}_{ijt} \equiv r_j \exp(\boldsymbol{x}^{\boldsymbol{v}\prime}_{\boldsymbol{ijt}}\boldsymbol{\omega})$, and baseline violations, $v^0_{i1t} \equiv \exp(\gamma_0 + \boldsymbol{f}^{\boldsymbol{v}\prime}_{\boldsymbol{it}}\boldsymbol{\gamma} + \eta_i)$, plus the shock ϵ^v_{ijt}. *Baseline violations* are the violations reported by the baseline certifier in an audit in the region of its headquarters, without accreditation inspection, apart from the shock ϵ^v_{ijt}. As baseline violations do not vary across certifiers, they capture ordinal differences in FMUs' levels of compliance. *Relative rigor* is the fraction of baseline violations that certifier j would report. As the model focuses on types and exogenous determinants of baseline violations and relative rigor, it describes a setting of adverse selection rather than moral hazard. Shock ϵ^v_{ijt} may create additional, random variation in violation reports. One the one hand, the shock may capture variation in rigor, for example, through auditors' concentration. On the other hand, it may also capture variation in compliance, for example through a shortage of FSC-compliant logging contractors.[44]

FMUs form expectations about their own compliance and certifiers' rigor, when deciding about participation in FSC certification and their certifier. Regarding compliance, I assume that FMUs predict the number of baseline violations conditional on their characteristics, $\mathbb{E}[v^0_{i1t}|\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}] = \exp(\boldsymbol{f}^{\boldsymbol{v}\prime}_{\boldsymbol{it}}\boldsymbol{\gamma})$. These are their *expected baseline violations*. FMUs do not explic-

[44] A certified forest manager in Canada, for example, reported a shortage of logging firms in her region, which reduces the forest managers' abilities to force the contractors to use FSC-compliant logging methods.

itly account for their unobserved compliance type η_i or the shock ϵ^v_{ijt} in their participation and certifier choice. These assumptions reflect the uncertainty reported by forest managers regarding their own compliance.[45] Nevertheless, η_i may correlate with a shock to the participation and certifier choice, as formalized in the next section. For example, part of forest managers' intrinsic motivation which does not correlate with their observed characteristics may affect not only their compliance, but also their preference for certain certifiers.

Regarding certifiers' relative rigor, FMUs' expectation is $\mathbb{E}[\tilde{r}_{ijt}|\overline{\boldsymbol{x}}^{\boldsymbol{v}}_{\boldsymbol{jt}}]$, where $\overline{\boldsymbol{x}}^{\boldsymbol{v}}_{\boldsymbol{jt}}$ is the certifier-market-level average of $\boldsymbol{x}^{\boldsymbol{v}}_{\boldsymbol{ijt}}$. That is, FMUs observe certifiers' rigor type, r_j, and the market's distance to the certifiers' headquarters, denoted as x^v_{1jt}. FMUs do not observe which audits will be inspected by the accreditation body, x^v_{2ijt}, but perfectly predict the likelihood of accreditation inspections for a given certifier in the market, i.e. the average $\overline{x}^v_{2jt}$.[46] Given these assumptions, the *expected relative rigor* is

$$ExpectedRigor_{jt} = r_j \exp(\omega_1 x^v_{1jt})\Big(1 + (\exp(\omega_2) - 1)\overline{x}^v_{2jt}\Big) \tag{4}$$

These assumptions capture insights shared by interviewed forest managers. On the one hand, they report that they ask managers of other certified FMUs about their experience with certifiers' stringency and expertise when choosing their certifier. On the other hand, they also report uncertainty, here captured by the shock ϵ^v_{ijt} and the indicator for accreditation inspections, x_{2ijt}.

4.2 Stage 2: Demand

Each forest management unit (FMU) i chooses whether to participate in FSC certification ($j > 0$) or not ($j = 0$). If FMUs choose FSC certification, they each have to select one available certifier in their market t, $j \in J_t$. Since each FMU makes these choices each year, a unique choice situation is defined by FMU i and market t. The FMU chooses the option

[45] Moreover, I could not consistently estimate η_i from the violations model. Attempts to account for $\exp(\eta_i)$ as random coefficients around $\mathbb{E}[v^0_{i1t}|\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}]$ were not successful, likely due to the limited degree of variation in the data.

[46] When FMUs choose certifiers, they do not know whether the accreditation body will inspect their audits by particular certifiers since the accreditation body needs to know the certificates issued by certifiers to assign inspections (ASI, 2019).

that maximizes their surplus u_{ijt}, indicated by y_{ijt}:

$$
\begin{aligned}
y_{ijt} &= \mathbf{1}(u_{ijt} \geq \max_{k \in J_t} u_{ikt}) \\
u_{ijt} &= \underbrace{\alpha_{it} p_{jt} + \beta^r_{it} ExpectedRigor_{jt} + \boldsymbol{x^{u\prime}_{jt}} \boldsymbol{\beta^x} + \boldsymbol{d'_{ijt}} \boldsymbol{\beta^d} + \boldsymbol{f^{u\prime}_{it}} \boldsymbol{\xi_j} + \xi_t + \Delta\xi_{jt}}_{V_{ijt}} + \epsilon^u_{ijt}
\end{aligned} \tag{5}
$$

The characteristics of main interest are p_{jt} and $ExpectedRigor_{jt}$. p_{jt} is the certification fee, $ExpectedRigor_{jt}$ is the expected relative rigor. β^r_{it}, thus, captures FMUs' preference or dislike of certifiers' expected relative rigor. $-\beta^r_{it}/\alpha_{it}$ quantifies the willingness to pay for leniency as $-\alpha_{it}$ captures the marginal utility of income.

The observable certifier characteristics $\boldsymbol{x^u_{jt}}$ are the certifiers' distance to their headquarters and an indicator for being available in a given market for the first year. Both characteristics may affect the certifiers' familiarity and local expertise in market t. $\boldsymbol{f^{u\prime}_{it}} \boldsymbol{\xi_j}$ are certifier fixed effects interacted with a constant and indicators for plantations and group certificates. They account for unobserved popularity of the certifier throughout the population and in those types of FMUs. $\Delta\xi_{jt}$ is an unobserved demand factor that may depend on exogenous shocks and certifier decisions such as advertising.

$\boldsymbol{d_{ijt}}$ is a vector of interactions of certifier and FMU characteristics. First, an indicator for participating in FSC certification for the first time accounts for entry cost. Second, I capture switching cost by including an indicator for having been certified by the same certifier in the previous year. I also include interactions of this indicator with the age of the certificate and an indicator for the end of a five-year cycle at which switching is less costly. Third, $\boldsymbol{d_{ijt}}$ includes an interaction of the number of years certifier j had been accredited in the first year in which FMU i was certified. That captures the importance of certifiers' experience and familiarity in FSC certification at the time of FMUs' entry. Fourth, $\boldsymbol{d_{ijt}}$ includes FMU characteristics interacted with a dummy for FSC certification. These interactions capture variation in certifier-invariant net benefits from participation across FMUs within the same market, while fixed effects ξ_t capture such variation across markets. Specifically, I allow for variation in those net benefits across FMUs with different clades of plants, countries' corruption perception indices and export values of wood chips.

The individual preference parameters α_{it} and β_{it} capture baseline preferences within the whole once-certified FMU population and heterogeneity in those preferences related to predicted FMU characteristics:

$$\begin{aligned}\alpha_{it} &= \bar{\alpha} + \tilde{\alpha} c_{it} \\ \beta^r_{it} &= \bar{\beta}^r + \tilde{\beta}^r_1 \mathbb{E}[v^0_{i1t}|\boldsymbol{f^v_{it}}] + \tilde{\beta}^r_2 \mathbf{1}\left\{\max\left(\mathbb{E}[v^0_{i1t}|\boldsymbol{f^v_{it}}] ExpectedRigor_{jt}\right)_{it} \geq 5\right\}\end{aligned} \tag{6}$$

c_{it} are predicted differences in cost, defined in the next section. $\mathbb{E}[v^0_{i1t}|\boldsymbol{f^v_{it}}]$ are FMUs' expected baseline violations, defined in Section 4.1. $\mathbf{1}\{\max(\mathbb{E}[v^0_{i1t}|\boldsymbol{f^v_{it}}] ExpectedRigor_{jt})_{it} \geq 5\}$ is an indicator for those choice situations in which FMUs expect the most rigorous certifier in their market to report at least five violations and, thus, to suspend their certificate. Using those predictions from the violations and pricing model to capture heterogeneity in preferences, instead of directly observable FMU characteristics allows for reduced dimensionality.

I normalize the mean utility of the outside option of not choosing FSC certification to zero: $u_{i0t} = \epsilon^u_{i0t}$. The taste shock ϵ^u_{ijt} follows an extreme value distribution according to a nested logit with two nests. The first nest is the set of FSC certifiers J_t, and the second is the outside option of no FSC certification, $j = 0$. The nest parameter λ measures the correlation of taste shocks for choosing an FSC certifier relative to choosing no FSC certification. Lower values of λ indicate higher substitutability among certifiers. $\lambda = 1$ implies independent taste shocks, translating into a multinomial logit model. The probability of choosing certifier j is the product of the probability of choosing FSC times that of choosing j conditional on having chosen FSC:

$$s_{ijt} \equiv s_{itFSC} s_{ijt|FSC} \tag{7}$$

where

$$s_{ijt|FSC} = \frac{\exp(V_{ijt}/\lambda)}{\sum_{k \in J_t} \exp(V_{ikt}/\lambda)}$$

and

$$s_{itFSC} = \frac{\left(\sum_{k \in J_t} \exp(V_{ikt}/\lambda)\right)^\lambda}{1 + \left(\sum_{k \in J_t} \exp(V_{ikt}/\lambda)\right)^\lambda}$$

These formulas show that the decision to participate in FSC certification depends on the characteristics of all available certifiers, including their expected relative rigor. Increasing minimum rigor levels or removing lenient certifiers from the choice sets, thus, inevitably affects participation in FSC certification. Estimating the model and simulating those changes allows assessing the size of the effects.

4.3 Stage 1: Pricing

I model certifiers pricing strategies to account for price effects in the counterfactual analysis of stricter accreditation. I build on the constant markup model (Train, 2009), which aligns with pricing practices in many industries (Shim and Sudit, 1995). That is, certifiers set markups in each market as a constant factor k_{jt} of the marginal cost c_{ijt}. I specify the markups as follows: the competitive fringe prices at cost ($k_{jt} = 1$), but all larger for-profit certifiers set k_{jt} to maximize profits, in best response to their competitors. Nonprofit certifiers price at a fixed, low margin. The model aligns with three features of many certification markets, including FSC. First, a small number of large certifiers dominate the market. Second, certifiers are differentiated. Third, the cost of certification varies by establishment (here an FMU) and market.

I add two more assumptions. First, I assume that the marginal cost c_{ijt} is the product of a certifier-market-level cost c_{jt} and an individual cost factor c_{it}. This aligns the pricing model with the demand model, where c_{it} is part of the price coefficient. Second, certifiers observe only the average of the individual cost factor per market, $\bar{c}_t = 1/N_t \sum_{i=1}^{N_t} c_{it}$, and market shares, rather than individual choice probabilities, when setting the markup. Each for-profit certifier j thus sets k_{jt} to maximize the expected profit based on that information:

$$\mathbb{E}[\pi_{jt}|c_{jt}, \bar{c}_t, s_{jt}, j \in J_t] = N_t(k_{jt} - 1)c_{jt}\bar{c}_t s_{jt}(\boldsymbol{k_t}) \tag{8}$$

where $\boldsymbol{k_t}$ is the vector of all markup factors in market t and $j \in J_t$ means that j is available in market t. Certifiers offer only a single version of certification for a given standard, which, here, is the FSC standard. The first-order condition is thus

$$(k_{jt} - 1)\mathbb{E}\left[\left(\frac{\partial s_{jt}(\boldsymbol{k_t})}{\partial k_{jt}}\right)\Big| c_{jt}, \bar{c}_t, s_{jt}\right] + s_{jt}(\boldsymbol{k_t}) = 0 \tag{9}$$

where

$$\mathbb{E}\left[\left(\frac{\partial s_{jt}(\boldsymbol{k_t})}{\partial k_{jt}}\right)\Big| c_{jt}, \bar{c}_t, s_{jt}\right] = \frac{\bar{\alpha} + \tilde{\alpha}\bar{c}_t}{\lambda} c_{jt} s_{jt}\left[1 - \lambda s_{jt} + (\lambda - 1)s_{jt|FSC}\right] \tag{10}$$

Overall, the price that an FMU i pays to certifier j is as follows:

$$p_{ijt} = p_{jt}c_{it} = k_{jt}c_{jt}c_{it} \tag{11}$$

The model departs from the standard Nash Bertrand pricing model in oligopolistic markets since certification costs vary across individual customers, here the FMUs. The standard model could account for such heterogeneous cost by modelling group-based price discrimination as in D'Haultfoeuille et al. (2019). However, the available data are not rich enough to estimate such a model.[47] The present model provides an alternative that imposes less sophistication on certifiers' pricing policies, but still captures profit-maximization when markups are set.

5 Estimation and results

The model is estimated sequentially, in reverse order.

5.1 Stage 3: Violation reporting

5.1.1 Estimation

To bring the model of violation reporting to the data, I add two assumptions. First, I assume that the unobserved compliance type is normally distributed in the population: $\eta_i | j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}} \sim N(0, \sigma_\eta)$. This implies that $\mathbb{E}[\exp(\eta_i)|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}] = \exp\left(\frac{\sigma_\eta^2}{2}\right)$. Second, I assume that the shock ϵ^v_{ijt} is independently and identically distributed for all certifiers and is exogenous to the determinants of relative rigor $\tilde{r}_{ijt}$ and baseline violations v^0_{i1t}.

I estimate the model using **Poisson pseudo-maximum-likelihood (PPML)**. To do so, I follow Santos Silva and Tenreyro (2006) by rewriting model (3) as

$$v_{ijt} = \exp(\mu^v_{ijt} + \eta_i + \tilde{\epsilon}^v_{ijt}) \tag{12}$$

where

$$\begin{aligned} \mu^v_{ijt} &\equiv log(r_j) + \boldsymbol{x^{v\prime}_{ijt}} \boldsymbol{\omega} + \boldsymbol{f'_{it}} \boldsymbol{\gamma} \\ \tilde{\epsilon}^v_{ijt} &\equiv \ln\left(1 + \frac{\epsilon^v_{ijt}}{\exp(\mu^v_{ijt} + \eta_i)}\right) \end{aligned} \tag{13}$$

[47] When splitting the FMUs into appropriate groups as in D'Haultfoeuille et al. (2019), the number of FMUs per group becomes so small that there is a large number of zero market shares.

This reformulation implies that $\mathbb{E}[\exp(\tilde{\epsilon}^v_{ijt})|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}, \eta_i] = 1$. A random sample of potential violation reports, v_{ijt}, would thus allow to estimate the model's parameters consistently by minimizing the following log-likelihood (Gourieroux et al., 1984; Wooldridge, 2010):[48]

$$\mathcal{L}^v = \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=1}^{J_t} \mathbf{1}\{v_{ijt} \text{ is observed}\} \left[v_{ijt} \log \left(\mathbb{E}[v_{ijt}|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}] \right) - \mathbb{E}[v_{ijt}|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}] \right] \quad (14)$$

where the normal distribution of η_i implies

$$\mathbb{E}[v_{ijt}|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}] = \exp\left(\mu^v_{ijt} + \frac{\sigma^2_\eta}{2} \right) \quad (15)$$

However, I do not observe a random sample of potential violation reports but the selected sample of violations reported by the certifiers FMUs have chosen. Estimating the model with this sample could lead to **sample selection bias**. In particular, less compliant FMUs might select less rigorous certifiers. These certifiers might, thus, end up reporting more violations than more rigorous certifiers.

To a substantial extent, the **observable controls** $\boldsymbol{f^v_{it}}$ account for such selection. I consider a large set of possible predictors derived from the variables in the FMU panel data and their interactions. I select the most relevant predictors by estimating model (12) with PPML using LASSO regularization with 5-fold cross-validation to avoid overfitting. I detail the procedure, including a few preliminary steps, in Appendix C.1.1.

Nevertheless, unobservable compliance factors, such as intrinsic motivation, might still affect violations and firms' certifier choices. In the model, the unobserved compliance type η_i captures such factors.[49] In the rest of this section, I formally derive the bias it introduces to account for it with a **control function**.

In the selected sample, the observed mean of violation reports does not correspond to the population mean from Equation (15). Instead, it conditions on FMUs' choices $\boldsymbol{y_{it}} \equiv [y_{i1t}, ..., y_{iJ_t t}]$ and their determinants $\boldsymbol{V_{it}} \equiv [V_{i1t}, ..., V_{iJ_t t}]$, defined by the demand model (5):

$$\mathbb{E}[v_{ijt}|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}, \boldsymbol{y_{it}}, \boldsymbol{V_{it}}] = \exp(\mu^v_{ijt}) \sum_{k \in \{1,...,J_t\}} y_{ikt} \, \mathbb{E}[\exp(\eta_i)|j, \boldsymbol{x^v_{ijt}}, \boldsymbol{f^v_{it}}, y_{ikt} = 1, \boldsymbol{V_{it}}] \quad (16)$$

[48]I estimate standard errors as being robust to heteroskedasticity and serial correlation. These standard errors correct for potential violations of the implicit PPML assumption of proportionality between the conditional mean and variance (Santos Silva and Tenreyro, 2006).

[49]The number of periods in the data is insufficient to consistently estimate η_i as a fixed effect.

$\mathbb{E}[\exp(\eta_i)|j, \boldsymbol{x_{ijt}^v}, \boldsymbol{f_{it}^v}, y_{ikt} = 1, \boldsymbol{V_{it}}]$ might vary across the chosen alternatives k and, thus, introduce sample selection bias. This is the case if η_i is correlated with the differences between the taste shocks $\epsilon_{i1t}^u - \epsilon_{ijt}^u, ...; \epsilon_{iJt}^u - \epsilon_{ijt}^u$ from model (5), even conditional on $\boldsymbol{f_{it}^v}$. Deriving the functional form of $\mathbb{E}[\exp(\eta_i)|j, \boldsymbol{x_{ijt}^v}, \boldsymbol{f_{it}^v}, y_{ikt} = 1, \boldsymbol{V_{it}}]$ allows to control for the bias.

The problem differs from the standard control function approach in two ways. First, the source of selection is a choice between multiple certifiers. It is a multinomial, not a binary choice. Second, the model is a count data model, so the conditional mean of $\exp(\eta_i)$ is not additive. To address these issues, I combine the approaches of Lee (1983) for multinomial selection problems and of Terza (1998) for count data models with binary selection. To my knowledge, I am the first to do so.

Following Lee (1983), I convert the problem of selection among J_t alternatives to J_t binary selection problems. To do so, I rewrite the selection problem from model (5) in terms of maximum order statistics:

$$y_{ijt} = 1 \text{ iff } V_{ijt} \geq e_{ijt} \text{ where } e_{ijt} \equiv max_{k \neq j}(V_{ikt} + \epsilon_{ikt}^u - \epsilon_{ijt}^u) \tag{17}$$

The marginal distribution of e_{ijt} is such that $F_j(V_{ijt}) = s_{ijt}$, the conditional choice probability defined in equation (7). e_{ijt} is transformed into a standard normal random variable by defining

$$e_{ijt}^* \equiv G_j(e_{ijt}) \equiv \Phi^{-1}(F_j(e_{ijt})) \tag{18}$$

e_{ijt}^* and η_i are thus jointly normally distributed with zero means, variances 1 and σ_η and correlation coefficient ρ_j, under the distributional assumption made on η_i. Since $G_j()$ is a strictly increasing function, (17) and (18) translate into $y_{ijt} = 1$ iff $e_{ijt}^* \leq G_j(V_{ijt})$.

As shown in Appendix C.1.2, I can then follow the steps suggested by Terza (1998) to derive the parametric form of the bias as

$$\mathbb{E}[\exp(\eta_i)|j, \boldsymbol{x_{ijt}^v}, \boldsymbol{f_{it}^v}, y_{ikt} = 1, \boldsymbol{V_{it}}] = exp\left(\frac{\sigma_\eta^2}{2}\right)\frac{\Phi(G_j(V_{ijt}) - \theta_j)}{\Phi(G_j(V_{ijt}))} \tag{19}$$

where $\theta_j \equiv \rho_j \sigma_\eta$. If $\theta_j = 0$, η_i and e_{ijt} are independent, then the control function $\frac{\Phi(G_j(V_{ijt})-\theta_j)}{\Phi(G_j(V_{ijt}))}$ equals 1 and is irrelevant. Otherwise, the estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{r}$ would be inconsistent when the control function is not included. A negative θ_j suggests that a higher random effect

η_i correlates with a lower e^*_{ijt}, i.e., with unobservable factors that make certifier j more attractive by reducing the value of the best alternative. In other words, a higher degree of non-compliance would correlate with choosing certifier j.

I estimate the model with the control function via PPML, as Egger et al. (2011) do when applying the approach of Terza (1998).[50] I obtain estimates of $G_j(V_{ijt})$ from the standard normal quantiles of the conditional choice probabilities estimated in a version of the demand model (5) that replaces expected relative rigor with some predictors of it.[51]

5.1.2 Identification

The previous sections have stated all the required assumptions to identify the rigor parameters r_j and $\boldsymbol{\omega}$. In this section, I provide intuition about the sources of their identification and implied assumptions. Ideally, one would identify r_j by comparing violation reports for pairs of audits, *it* and $i't'$, conducted by different certifiers, j and 1, but are otherwise equivalent. "Otherwise equivalent" means that they would result in the same number of violation reports if conducted by the same certifier. Similarly, one would identify ω_1 by comparing audits with different distances to the certifiers' headquarters that are otherwise equivalent. The identification of ω_2 would rely on comparing otherwise-equivalent audits with and without accreditation inspection. In practice, one does not need perfectly equivalent pairs for direct comparison but can exploit the model's functional form to use any variation in the degree of equivalence.

The challenge is that I do not observe but need to predict the degree of equivalence. The key assumptions to identify certifiers' relative rigor are, thus, exogeneity of (i) the certifier identity, (ii) the market's distance to certifiers' headquarters, and (iii) accreditation inspections to unobserved determinants of violation reports. That is exogeneity to unobserved factors that render two audits nonequivalent. That exogeneity assumption is needed to interpret expected violation reports as expected changes in violation corrections and, consequently, quality in the counterfactual analysis. Apart from that, the paper's main results rather depend on identifying *expected* relative rigor.[52] Its identification only requires exogeneity to

[50] In Egger et al. (2011), selection is binary. These authors can, thus, apply the control function suggested by Terza (1998) without combining it with Lee (1983) as I do.

[51] I estimate a simplified version of the demand model (5), but without including expected relative rigor, baseline violations, and prices explicitly. Certifier-market constants capture variation in expected relative rigor across regions, years, and certifiers. Interactions of certifier dummy variables with observed FMU characteristics capture variation in baseline violations.

[52] Expected relative rigor drives participation and certifier choices and, thus, the willingness to pay for

unobserved determinants of violation reports which FMUs *expect*. In the rest of the section, I add references to that type of exogeneity in brackets.

Exogeneity of the certifier identity requires that the rates at which the determinants of violation reports $\boldsymbol{f}_{it}$ affect rigor do not differ systematically across certifiers (in FMUs' expectation). That is, their effect on (expected) violation reports, $\boldsymbol{\gamma}$, must be certifier-invariant. Apart from threats in this regard, the certifier identity would only be endogenous if the observables $\boldsymbol{f}_{it}^{v}$ did not capture all joint determinants of FMUs' certifier demand and compliance and if the control function was not well identified at the same time. The control function can be identified based on parametric form assumptions or exclusion restrictions, i.e., determinants of demand that do not affect violation reports. The certifier-market-level constants δ_{jt} in the demand model capture variation across markets that cannot stem only from determinants of violation reports. This variation, thus, satisfies an exclusion restriction. However, within markets, all observed determinants of individual FMUs' demand may also affect violation reports. Therefore, I need to rely on the distributional form assumptions to control for potential within-market selection on unobservables.

Regarding the exogeneity of markets' **distance to certifiers' headquarters**, there is only one negligible concern: Most certifiers have their headquarters in Europe, where (expected) compliance tends to be higher than in other continents. Observable controls account for this across-market variation in compliance.

Accreditation inspections are assigned based on indicators of certifiers' leniency and FMUs' compliance, market shares of certifiers, practical considerations such as the location of assessors, and some random variation.[53] The model controls for many determinants of leniency and compliance and for potential correlation of market shares with expected violation reports, as I discuss in Appendix C.1.3. Conditional exogeneity of the (expected) inspection assignments requires that the accreditation body does not assign inspections based on additional information about leniency or compliance (if FMUs anticipate that). Remaining variation in market shares, practical considerations, and random variation allow then to identify ω_2.

leniency and changes in choices in the counterfactual simulations of stricter accreditation.

[53]This information is based on ASI (2021) and an interview with the accreditation body's staff.

5.1.3 Results

Table 5 presents the estimates of the main parameters. Columns (1) and (2) show the results of the estimation without and with control functions, respectively. Most of the coefficients θ_j from the control functions are statistically different from zero. As outlined above, this finding suggests that the estimates from Column (1) are inconsistent. In the following, I, thus, interpret the results from Column (2). I include standard errors that are robust to heteroskedasticity and serial correlation, but do not account for the use of generated regressors in the control functions.[54]

Table 5: No. of violation reports (by year)

	Without control function (1)	*With control function* (2)
Log(r_j), rigor type:		
Certifier 2	−0.291 (0.246)	−0.442*** (0.135)
Certifier 3	0.350* (0.201)	0.241*** (0.081)
Certifier 4	−0.053 (0.160)	−0.076 (0.112)
Certifier 5	−0.509*** (0.172)	−0.503*** (0.075)
Certifier 6	0.531*** (0.140)	0.429*** (0.054)
Certifier 7	−0.875*** (0.226)	−0.584*** (0.125)
Small certifiers	−0.165 (0.246)	0.024 (0.168)
$\boldsymbol{\omega}$*:*		
Audit inspected by accreditation body	0.776*** (0.156)	0.793*** (0.054)
Average distance to certifier's headquarter by market in 1000 km	−0.034** (0.017)	−0.033*** (0.007)
θ_j from control functions:		
Certifier 1		−0.061 (0.040)
Certifier 2		−0.256*** (0.091)
Certifier 3		−0.222*** (0.057)
Certifier 4		−0.083 (0.064)
Certifier 5		−0.080 (0.054)
Certifier 6		−0.228*** (0.033)
Certifier 7		0.545*** (0.174)
Small certifiers		0.270 (0.219)
Controls $\boldsymbol{f_{it}}$	Yes	Yes
Observations (choice situations)	3,810	3,810

Notes: Standard errors are robust to heteroskedasticity and serial correlation. They do not account for the variance of the generated regressors used as the control function, yet.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Rigor types r_j differ significantly, both statistically and economically, across certifiers. Figure 3 (a) ranks certifiers by plotting the estimates of r_j with 95% confidence intervals. The lowest

[54]Bootstrapped standard errors allow me to account for the use of generated regressors. However, there is little guidance on bootstrapping and interpreting results from bootstrap samples in the estimation of a model with multiple stages using data that cover the whole population, as in this paper. For example, I decided to use stratified sampling with replacement to ensure that all existing certifier-market combinations are included. Estimates from a preliminary set of 550 such bootstrap samples indicate that the statistical uncertainty is larger than the standard errors presented here suggest. Nevertheless, a share of the coefficients remains significant at a 10%-level, particularly three out of the seven estimated rigor types and the coefficient on accreditation inspections, which is helpful for identification in the choice model.

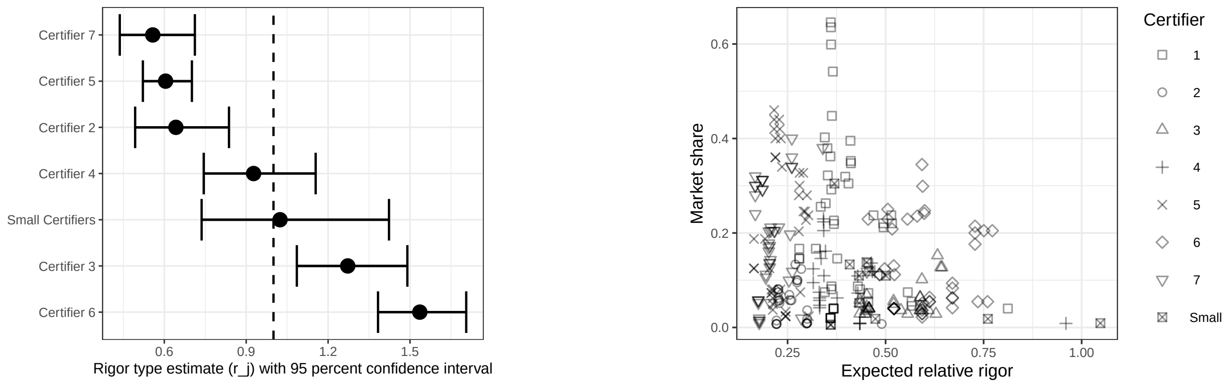
ranked certifier, Certifier 7, reports only 56% of the number of violations the baseline certifier reports, everything else equal. Apart from the small certifiers, the best certifier is expected to report 54% more than the baseline certifier. Two facts support the identification of the most lenient certifiers: suggested by the estimates: First, the only two certifiers that FSC's accreditation body partially suspended from forest management certification in 2011-2020 appear among the three lowest ranked rigor types. Second, NGOs such as Earthsight (2020) have reported Certifier 7's failure to find serious violations such as illegal logging.

Why should certifiers have different rigor types? Different ownership types and intrinsic motivations of managers or staff might explain such variation. However, different rigor types could appear and survive even if the rigor type was a strategic long-run choice of for-profit certifiers without intrinsic motivation. The costs and benefits from increasing rigor may vary across certifiers. Costs may include increased audit costs and reduced revenues if FMUs prefer leniency. Benefits arise from a reduced likelihood of losing accreditation and market access. Classic reputation theory suggests that larger certifiers with higher profits have reason to be more concerned about losing market access and may, thus, be more rigorous than smaller certifiers (Klein and Leffler, 1981). However, such a pattern is not found here. Certifiers 2 and 7 are very large certifiers both in FSC and other markets. They are the only two publicly traded FSC certifiers. The overall correlation of certification companies' total revenue in 2020 with the relative rigor estimates is negative and statistically significant. Given the low number of certifiers, this finding cannot be interpreted as a pattern in the opposite direction, but, rather, as motivation for further research. An effect of rigor on the likelihood of accreditation suspensions can still rationalize the heterogeneity of rigor if certifiers differ in the degrees to which they discount future profits. Appendix C.1.5 discusses all those potential explanations of rigor heterogeneity in more detail.

Rigor also varies across accreditation inspections and distance to the certifier's headquarters. The estimates suggest that certifiers report on average 121% more violations in audits with accreditation inspections than in equivalent audits without inspections. Potential explanations for this are moral hazard in uninspected audits or extra rigor in inspected audits due to a behavioral effect, also called Hawthorne effect (Landsberger, 1958). On average, certifiers report 29% fewer violations when they audit 10,000 km farther away from their headquarters, potentially due to reduced quality control by headquarters.

Nevertheless, the certifier-invariant predictors $\boldsymbol{f}_{it}^{v}$ explain much more of the variation in violation reports than differences in rigor. Compared to using the mean number of violation reports, controlling for $\boldsymbol{f}_{it}^{v}$ reduces the cross-validated prediction error by 11%, while the

Figure 3: Relative rigor predictions



(a) Rigor type

(b) Expected relative rigor and market shares by certifier (The darker a shape, the more observations it covers.)

certifier variables r_j and $\boldsymbol{x}^{\boldsymbol{v}}_{\boldsymbol{ijt}}$ only reduce it by 2 additional percentage points (and 2.7% if one excludes the predictors $\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}$). The in-sample R-Squared, adapted for the count data model following Cameron and Windmeijer (1996), is 0.28 for the full model. That suggests that the included variables are predictive of violation reports, but that a substantial share of the variation remains unexplained. Nevertheless, recall that the main purpose of the model is not to predict realized violation reports, but FMUs' expectations of relative rigor and violation reports (next to a rigor ranking that allows to simulate the suspension of the most lenient certifiers).

Table 6: Summary statistics of the predictions about violation reporting

Statistic	N	Median	Mean	St. Dev.	Min	Max
Expected violation reports	34,794	0.40	0.61	0.73	0.02	17.23
Expected relative rigor $ExpectedRigor_{jt}$	34,794	0.36	0.39	0.16	0.16	1.05
Expected baseline violations $\mathbb{E}[v^0_{i1t}\|\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}]$	34,794	1.08	1.64	1.81	0.09	25.71

Table 6 provides summary statistics of the model's predictions of FMUs' expectations, namely the expected relative rigor $ExpectedRigor_{jt}$, expected baseline violations $\mathbb{E}[v^0_{i1t}|\boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}]$ and their product, expected violation reports. Unsurprisingly, expected violation reports have about the same mean, but a slightly smaller range than realized violation reports. The predictions of expected relative rigor suggest that FMUs expect certifiers to report only about 40% of the baseline violations, on average. But that share ranges from 16% to 105%. Figure 3 (b) plots those shares against the market share of each certifier in each market. The plot shows substantial variation in expected relative rigor across certifiers and across markets within certifiers. In the next section, I will use the variation across markets to identify FMUs' dislike of rigor.

5.2 Stage 2: Demand

5.2.1 Estimation and identification

I estimate the demand model in two steps. Following Berry (1994), I split the utility into certifier-market-level utility, δ_{jt}, and within-market variation, $\tilde{V}_{ijt}$:[55]

$$
\begin{aligned}
u_{ijt} =& \delta_{jt} + \tilde{V}_{ijt} + \epsilon^{u}_{ijt} && (20) \\
\delta_{jt} =& \bar{\alpha} p_{jt} + \bar{\beta}^{r} ExpectedRigor_{jt} + \boldsymbol{x}^{\boldsymbol{u}\prime}_{\boldsymbol{jt}} \boldsymbol{\beta}^{\boldsymbol{x}} + \xi_j + \xi_t + \Delta\xi_{jt} && (21) \\
\tilde{V}_{ijt} =& \tilde{\alpha} c_{it} p_{jt} + \\
& \left(\tilde{\beta}^{r}_{1} \mathbb{E}[v^{0}_{i1t} | \boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}] + \tilde{\beta}^{r}_{2} \mathbf{1} \left\{ \max \left(\mathbb{E}[v^{0}_{i1t} | \boldsymbol{f}^{\boldsymbol{v}}_{\boldsymbol{it}}] ExpectedRigor_{jt} \right)_{it} \geq 5 \right\} \right) ExpectedRigor_{jt} + \\
& \boldsymbol{d}^{\prime}_{\boldsymbol{ijt}} \boldsymbol{\beta}^{\boldsymbol{d}} + \boldsymbol{f}^{\prime}_{\boldsymbol{it}} \boldsymbol{\xi}_{\boldsymbol{j}} && (22)
\end{aligned}
$$

First, I estimate the **nested logit** model (20) with constants δ_{jt} and the determinants of $\tilde{V}_{ijt}$ from (22). I maximize the following log-likelihood:

$$
\mathcal{L}^{u}(\boldsymbol{\Lambda}) = \sum_{t=1}^{T} \sum_{i=1}^{N_t} \sum_{j=1}^{J_T} y_{ijt} \log s_{ijt}(\boldsymbol{\Lambda}) \tag{23}
$$

where $\boldsymbol{\Lambda}$ is the vector of parameters $\tilde{\alpha}, \tilde{\boldsymbol{\beta}}^{\boldsymbol{r}}, \boldsymbol{\beta}^{\boldsymbol{d}}, \boldsymbol{\xi}_{\boldsymbol{j}}, \boldsymbol{\delta}, \lambda$ and $s_{ijt}(\boldsymbol{\Lambda})$ the nested logit choice probability defined in equation (7).

Second, I estimate model (21) in a pooled **two-stage least squares (2SLS)** regression. Prices are likely endogenous since the structural error, $\Delta\xi_{jt}$, captures unobserved popularity and may thus affect market shares and hence the markup. Expected relative rigor may also be endogenous since market shares can affect the likelihood of accreditation inspections $\bar{x}_{2jt}$ and since accreditation inspections increase expected relative rigor. Following Gandhi and Houde (2019), I use differentiation instruments to account for the endogeneity of both variables; namely,

[55] This approach seems more reliable than a control function approach, given the difficulties with the compatibility of the latter and a supply model based on profit maximization as discussed in Train (2009).

1. The number of close-by rivals, $\sum_{k \neq j} 1(|x_{kt} - x_{jt}| < \kappa)^2$, where κ is the standard deviation of x_{kt} across all markets
2. The interaction of that instrument with some mean demographic by market, $\bar{f}_t$

Such instruments are exogeneous for those characteristics x_{jt} for which the utility model controls. Any effect of differentiation in such characteristics on market shares goes through the differences in FMUs' utility levels across certifiers. In the main specification, I construct such instruments for certifiers' experience as x_{jt}, i.e., the number of years they have been FSC-certifiers. The model's combination of certifier and market fixed effects perfectly captures certifiers' experience. I use the average longitude in the market as the mean demographic $\bar{f}_t$.

The instruments are relevant since certifiers with more close-by rivals tend to have lower market shares, everything else equal. The demographic accounts for variation in the effect of close-by rivals on market shares across markets. For prices, lower market shares lead to lower markups. The effect of market shares on expected relative rigor goes through the likelihood of accreditation inspections. The accreditation body generally tends to inspect larger certifiers more often than smaller ones. In some cases, increased competition can also lead to more frequent inspections, especially for relatively new certifiers in a market.

There is one potential concern with the abovementioned phenomenon: The accreditation body may inspect more in markets with more competition because they expect rigor to decrease with increased competition. If this were the case, I might misinterpret the coefficient on expected relative rigor since I do not control for variation in rigor within certifiers across markets. I can test for this by examining the association of the instruments with the variation in violation reports that the predictor variables used above do not explain. To do so, I regress the residual of violation reports $\hat{\epsilon}^v_{ijt}$ from model (3) on the differentiation instruments. The coefficients on all instruments are far from significant. Hence, variation in unmeasured rigor does not seem to drive the variation induced the instruments by in accreditation inspections and $ExpectedRigor_{jt}$. This is true for the rigor that affects the number of violation reports. Instead, the accreditation body may inspect more in cases of increased competition as they can rely less on historical information about auditors in such contexts. This explanation supports the assumption of instrument exogeneity.

Identification of the price and rigor coefficients, $\bar{\alpha}$ and $\bar{\beta}^r$, relies on variation in prices and rigor across markets within each certifier, following different patterns for different certifiers. Identification requires such variation since fixed effects absorb across-certifier variation that is the same in all markets and across-market variation that is the same for all certifiers. Certifier fixed effects are important since they control for unobserved differences in certifiers' general expertise and efficiency which might correlate with rigor. Market fixed effects matter as FMUs' net benefits of FSC certification are probably sensitive to market-specific changes in demand for certified wood. Replacing market fixed effects with separate region and year fixed effects may omit some of these changes but may allow for the use of additional variation for the identification of $\bar{\alpha}$ and $\bar{\beta}^r$. Appendix-Table A11 shows that such replacement would not affect the results much.

The identifying variation for the rigor coefficient $\bar{\beta}^r$, stems mainly from the variation in the likelihood of accreditation inspections since the 2SLS-model (21) controls for the distance to certifiers' headquarters. Appendix-Figure A20 shows differential variation in the likelihood of accreditation inspections across markets within certifiers. The choice of relevant instruments ensures that the instruments predict sufficient exogenous variation in that likelihood. Appendix C.2.3 presents tests for that.

FMUs' net benefits from participation in FSC certification are identified based on the variation in the timing of FMUs' decisions to join or leave FSC. The identification exploits variation across markets with different choice sets of certifiers and different certifier characteristics. In addition, it uses variation across FMUs with different characteristics within the same market to account for heterogeneity in their benefits.

5.2.2 Results

Table 7 presents the parameter estimates from the certifier-market-level utility model (21). Column (1) presents the OLS results. The 2SLS results from Column (2) are the main specification since I reject the hypothesis of consistent OLS results based on the Wu-Hausman test. Table 8 presents the parameter estimates that capture heterogeneity in FMUs' preferences within markets from the nested logit model (22). In both tables, I report heteroskedasticity and serial correlation robust standard errors. They do not yet account for the use of generated regressors, which most likely increases the statistical uncertainty substantially.[56] I present reassuring tests of instruments' relevance in Appendix C.2.3. Tables A8 and A9 in the Appendix show the coefficient estimates from the first stages.

[56] Estimates from the preliminary set of 550 bootstrap samples suggest statistical insignificance for many of the parameters of interest. However, the 95% confidence interval of the average willingness to pay for leniency is entirely positive.

Table 7: Selected preference estimates at the certifier-market-level

	OLS	*2SLS*
	(1)	(2)
$\bar{\alpha}$: Price at certifier-market-level in 1K USD	-0.069^{***} (0.003)	-0.770^{***} (0.178)
$\bar{\beta}^r$: Expected relative rigor	-1.367^{***} (0.046)	-17.366^{***} (1.132)
$\boldsymbol{\beta^x}$:		
First year certifier is available	0.294^{***} (0.019)	1.698^{***} (0.195)
Average distance to certifier's headquarter by market in 1000 km	-0.025^{***} (0.001)	-0.139^{***} (0.016)
Certifier FE	Yes	Yes
Market FE	Yes	Yes
Observations	34,794	34,794
Adjusted R^2	0.649	-1.006
Residual Std. Error (df = 34733)	0.484	1.156
F Statistic	$1{,}072.535^{***}$ (df = 60; 34733)	
Wu-Hausman stat.		723.996*** (df=2,34731)
Weak IV stat. (Expected relative rigor)		343.283*** (df=2,34733)
Weak IV stat. (Price at certifier-market-level in 1K USD)		77.239*** (df=2,34733)

Notes: Heteroscedasticity robust standard errors. Standard errors are not yet corrected for the use of generated regressors and regressands. The dependent variable are the estimates of mean utility by certifier and market, obtained from MLE of the certifier-market constants in the nested logit model of FMUs' certifier choice. The outside option in the nested logit includes both dropping out of FSC certification or not getting certified yet. Dropping out is permanent in the data, such that only the first year after drop-out is included. In the 2SLS regression here, the outside option is excluded since its mean utility is normalized to zero for all markets in the nested logit.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The main result is as expected: the estimate of $\bar{\beta}^r$, the coefficient on certifiers' expected relative rigor, is negative. This sign suggests that, all else being equal, FMUs prefer a certifier whom they expect to report fewer violations than other certifiers. That is, they prefer a more lenient certifier. The coefficients $\boldsymbol{\tilde{\beta}^r}$ capture heterogeneity in this preference across FMUs with different levels of compliance. Surprisingly, the point estimates suggests that FMUs with more baseline violations, a proxy for lower compliance, dislike rigor slightly less. This seems to be even more true for FMUs that can expect the most rigorous certifier to find at least five violations, i.e., that risk losing their certificates. Yet, these two coefficients are not significant. In particular, the coefficient on the indicator for more than five expected violation reports is very noisy, given the small number of observations in such a situation.

The difference between Columns (1) and (2) in Table 7 suggests that the coefficients on prices and expected relative rigor are underestimated in absolute terms if I do not account for their endogeneity. Such underestimation is expected: larger demand shocks at the certifier-market-level should correlate with higher markups, and thus, the OLS results underestimate FMUs' dislike of higher markups. Larger demand shocks should also correlate with more frequent accreditation inspections, which are associated with more violation reporting. Thus, the OLS results underestimate FMUs' dislike of rigor and violation reporting when not accounting for the endogeneity of accreditation inspections.

A few other parameter estimates are worth highlighting. First, the baseline price coefficient $\bar{\alpha}$ is negative, as expected. Second, FMUs with larger cost factors, i.e., larger total prices

Table 8: Selected estimates of heterogeneity in preferences across FMU types (Nested Logit)

$\tilde{\alpha}$: Price at certifier-market-level in 1K USD X ...	
Cost factor	0.020*** (0.008)
$\tilde{\boldsymbol{\beta}}^r$: Expected relative rigor X ...	
Expected baseline violations	0.082 (0.097)
≥ 5 violations reports expected from most rigorous certifier	0.519 (1.290)
Forest chose same certifier last year	2.496*** (0.345)
Forest chose same certifier last year x Yrs. with FSC cert.	0.051*** (0.014)
Forest chose same certifier last year x Recertification year x FSC	−0.076 (0.166)
FSC Certifier has office in forest's country	0.549*** (0.141)
Is first FSC certifier (entry cost)	−2.457*** (0.281)
λ:	
Within FSC nest correlation	0.464*** (0.063)
Certifier-market FE	Yes
Group cert. indicator x Certifier FE	Yes
Plantation indicator x Certifier FE	Yes
FMU characteristics x FSC indicator	Yes
Observations (choice situations)	6,250
Log Likelihood	−2,744.536

Notes: MLE of nested logit choice model with R package mlogit. hteteroscedasticity and serial correlation robust standard errors. Standard errors are not yet corrected for the use of generated regressors. The outside option includes both dropping out of FSC certification or not getting certified yet. Dropping out is permanent in the data, such that only the first year after drop-out is included. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

for their forest area, are less price sensitive. This association may seem counterintuitive. However, this is plausible since FMUs with larger cost factors often have higher revenues, for example, since they tend to have more extensive forests. Third, the net benefit from FSC certification is substantially lower in the first year of participation. This finding reflects a one-time cost of "entering" the certification system. Fourth, FMUs are more likely to stay with their previous certifiers, suggesting the presence of switching costs. Fifth, certifiers are more popular closer to their headquarters and in countries where they have an office, ceteris paribus. Finally, the estimated correlation of taste shocks within the FSC nest $1 - \lambda$ is far from perfect correlation but also substantially different from independence. This finding suggests that certifiers are not perfect substitutes but also not completely different in terms of their yearly unobservable characteristics, captured by the taste shocks.

Shopping for leniency: in the rest of this section, I analyze the extent of FMUs' preference for leniency and their willingness to pay for it. Table 9 presents summary statistics of the elasticities of individual market shares with respect to changes in certifier-market-level prices and expected relative rigor. Recall that expected relative rigor is the portion of baseline

Table 9: Price and rigor elasticity estimates

Statistic	Median	Mean	Min	Max
Price at certifier-market-level	−4	−4	−19	−0
Expected relative rigor	−11	−12	−39	−0

Notes: Percentage point change in conditional choice probabilities per percentage change in certifier-market-level prices and expected relative rigor

violations that FMUs' expect a certifier to report. On average, a 1% increase in this portion is associated with an average decrease in individual market share of 12%. This sizable elasticity is more than three times the average own-price elasticity.

Table 10: Estimated willingness to pay

Statistic in 1000 USD for a certifier that ..	Median	Mean	Min	Max
Reports 1 SD lower fraction of violations	4.08	4.12	1.50	10.01
Reports 1 violation less	22.56	34.44	0.99	262.50
Has a 1 SD closer headquarter	0.52	0.52	0.19	1.09
Is last year's certifier	3.46	3.52	3.26	10.06
Has an office in your country	0.76	0.78	0.72	2.21
Skipping the first year with FSC (entry cost)	3.41	3.47	3.21	9.90

Notes: The numbers correspond to the coefficient estimates divided by the price coefficient, multiplied by the within-market standard deviation across certifiers for the numeric variables in rows 1 and 3, and divided by the number of baseline violations in row 2.

Table 10 shows the estimated willingness to pay (WTP) for leniency and other certifier characteristics. My principal measure of the WTP for leniency is the WTP for a certifier that reports a one standard deviation lower fraction of baseline violations than its market competitors, i.e., a certifier that is one standard deviation more lenient than other certifiers in the same market. On average, that WTP is 4123 USD. The average standard deviation in leniency is 0.2, i.e., a 20-percentage-point difference in the reported fraction of baseline violations. The WTP for a certifier that reports one violation less is accordingly higher. However, as I discuss in the next section, the WTP of 4123 USD is already substantial. It is roughly as large as the estimated "entry cost" in the first year of certification and more than half the average predicted certification fee. At the same time, the WTP for other characteristics may limit the degree to which FMUs shop for more lenient certifiers. Most importantly, remaining with the same certifier as in the previous year is of similar value as leniency. This implies that switching to another certifier in years when the previous certifier reports fewer violations due to fewer accreditation inspections would be costly. For most FMUs, their preference for leniency is, thus, relevant primarily for their initial participation and certifier choice or when a suspension of their certifier's accreditation forces them to choose a new one, as the counterfactual analysis will simulate.

5.2.3 Discussion

The WTP for leniency captures the net opportunity cost of violation reports for FMUs to the extent that FMUs consider these costs when choosing a certifier. These net costs may include economic private net costs as well as societal benefits if FMUs care about these benefits.

Is the WTP a reasonable estimate of violation reports' economic private net costs? There is no estimate of these costs in the literature so far. However, the discussion in Section 2.3 suggests that the economic private cost of violation correction is of the same order of magnitude as the estimated WTP for a certifier that reports one less violation.[57]

Table 11: Estimated benefits and costs of FSC certification

Statistic in 1000 USD	Median	Mean	Min	Max
Net benefits, gross of fees and violation correction cost	20.80	20.57	−2.01	41.13
Violation correction cost	8.52	9.29	3.73	25.59
Value of certification fee	3.14	3.24	1.27	11.67
Net benefit, net of fees and violation correction cost	9.03	8.04	−10.73	20.01
Potential additional revenues due to FSC	255.79	1,887.37	0.08	93,122.94

Notes: The numbers are computed according to the model estimates for FMUs' chosen FSC-certifier. The cost of violation correction are the WTP for one less violation report multiplied by the number of violations reported by the chosen certifier. The value of the certification fee is the market-level certification fee since the cost factor is part of the price coefficient. I compute the potential additional revenues by assuming a price premium of 5%, half the median premium from choice experiments, the average price per cubic meter of roundwood imports in the member states of the UNECE in 2017, and the average number of cubic meters produced in FSC-certified forests per hectare in 2017.

How does the perceived cost of violation reports relate to benefits from FSC certification? Table 11 summarizes the model's predictions of the benefits and costs of FSC certification, focusing on the chosen certifier for each FMU. The first row shows the net benefits if there were neither fees nor expected violation reports. The next row presents the opportunity cost of the expected violation reports by the chosen certifier, computed based on the WTP

[57]For example, if 60% of violations induce a cost of 1000 USD, 30% a cost of 10,000 USD, and 10% a cost of 100,000 USD, then the average cost is 13600 USD. Such a distribution corresponds roughly to the qualified guess of violation correction cost in a small, randomly drawn sample, as described in Section 2.3.

for leniency. At the median, this cost corresponds to almost half of the net benefits in the first row. The fourth row of Table 11 shows the net benefits after subtracting the violation correction cost and the less substantial value of the certification fee (in the third row) from the values in the first row. The remaining net benefits are only 9,000 USD at the median, with many negative values in the distribution. This finding suggests that many FMUs are very sensitive to changes in available rigor levels when deciding whether to participate in FSC certification. The counterfactual analysis investigates this sensitivity further.

How do these net benefits and the WTP for leniency compare to the gross benefits from certification, such as additional revenues? The parameter estimates do not allow me to compute the increase in revenue or other gross benefits through FSC certification since most coefficients can capture both cost and benefits. The last row of Table 11 summarizes back-of-the-envelope estimates of additional revenue from FSC certification based on the median willingness to pay a premium, as reported in choice experiments in the 2000s (Cai and Aguilar, 2013). At the median, the additional revenue estimates exceed the net benefits by a factor of more than one hundred. What can explain this gap? On the one hand, the additional revenue estimates are likely an upper bound. The willingness to pay a premium stated in choice experiments often exaggerates a consumer's willingness to pay during purchases. Anecdotal evidence further suggests that price premia for certified wood have decreased since the choice experiments covered by Cai and Aguilar (2013). On the other hand, parameters other than the violation correction cost can account for those certification costs not captured by the estimated violation correction cost and certification fees. The estimated violation correction cost is based on the WTP for leniency in terms of the number of violation reports, while other dimensions of leniency can also impact FMUs' cost. For example, some certifiers may report violations that are less expensive to correct than others. In addition, it matters how easily certifiers accept corrections of violations in follow-up audits, information that is not included in the data thus far.

5.3 Stage 1: Pricing

I use the estimated demand parameters to compute certifiers' markups and marginal cost computed according to the price model. I do so to account for price effects in the counterfactuals of stricter accreditation. To compute markups, I replace c_{jt} in derivative (10) with

$\frac{p_{jt}}{k_{jt}}$. The profit-maximizing markup factor is then

$$k_{jt} = \frac{\frac{\bar{\alpha}+\tilde{\alpha}\bar{\tilde{c}}_{it}}{\lambda} p_{jt} \left[1 - \lambda s_{jt} + (\lambda - 1) s_{jt|FSC}\right]}{\frac{\bar{\alpha}+\tilde{\alpha}\bar{\tilde{c}}_{it}}{\lambda} p_{jt} \left[1 - \lambda s_{jt} + (\lambda - 1) s_{jt|FSC}\right] + 1}$$

where I replace the coefficients and variables with the predictions from the model. Table 12 presents the markup and marginal cost estimates. Mark-ups constitute 33% of the market price on average, with most observations being below that level.

Table 12: Mark-up and marginal cost estimates

Statistic	N	Median	Mean	Min	Max
Prices (1000 USD)	34,794	7.49	9.44	0.31	128.29
Marginal cost (1000 USD)	34,794	6.22	8.08	0.21	110.38
Mark-ups (percentage of cost)	34,794	30.00	33.00	6.79	135.53

6 Counterfactual Analysis

I use the estimated model to implement two sets of counterfactual exercises. They investigate whether increasing the minimum level of certifiers' rigor can increase the number of violation reports, hence, corrections of violations and, thus, quality among all once-certified FMUs. In particular, I analyze to what extent effects on participation counteract positive effects on quality among certified FMUs. I also attempt to quantify welfare effects.

In the first set of counterfactuals, I consider direct shifts in the global minimum level of rigor. This exercise explores the effects label owners such as FSC might achieve if they had direct control over minimum rigor. It also allows to illustrate how effects can develop from very small to very large increases of minimum rigor. Finally, this exercise quantifies how costly rigor increases are for certifiers due to FMUs' willingness to pay for leniency.

In the second set of counterfactual scenarios, I draw attention to the enforcement of increased minimum rigor. For many label owners, including FSC, credible threats of suspending market access are arguably the only way to control certifiers' rigor. I, therefore, simulate the suspension of lenient certifiers' accreditation. I then isolate the effects of accreditation suspension due to changes in rigor from the effects due to changed choice sets. To do so, I consider equivalent shifts in those certifiers' rigor. Comparison to these isolated effects allows for discussion of some drawbacks of certification systems with external certifiers.

In all cases, I simulate changes in quality, participation, and agents' surplus. I account for mechanical effects of increased rigor and indirect effects through changes in certification choices and prices. I keep everything else constant and discuss the implications after presenting the results.

6.1 Procedure

For each scenario, I solve for the new market equilibrium under the counterfactual assumptions. I allow the FMUs to adjust their certification choices and the certifiers their markups and prices. Specifically, I carry out the following steps:

First, I implement the counterfactual assumptions for 2019, the last year of my data. The observed market structure in 2019 is the baseline scenario.

1. For the first scenario, I consider 10% increases from 0 to 500% in the minimum level of expected relative rigor across all certifiers and markets, $\min\{rigor_{jt} \forall j \in J, t \in T\}$. I call this level the *global minimum*. For each new global minimum, I increase the expected relative rigor to this level wherever it is below at baseline.
2. I consider three cases for the second scenario: suspension of the most lenient, two most lenient, or three most lenient certifiers' accreditation.
   i. I rank certifiers according to their constant rigor type, r_j.[58] I remove the suspended certifiers from FMUs' choice sets J_t.
   ii. For each set of suspended certifiers, I also conduct a separate simulation where I do not remove them from the choice sets but shift their expected relative rigor to the level of the next most lenient certifier in each market. The lowest available level of expected relative rigor in each market is then the same as in the case of suspension, but the set of available certifiers is the same as in the baseline scenario.[59]

Second, I solve for the new set of choice probabilities $s_{ijt}(\boldsymbol{k}'_t)$ and markups $\boldsymbol{k}'_t$ based on the parameter estimates from Table 7, Column (2), and Table 8. All other variables and

[58]The ranking according to each certifier's global minimum level of expected relative rigor is the same.

[59]These rigor shifts differ from the shifts in the global minimum analyzed above: the lowest ranked certifiers' leniency varies from market to market, such that their global suspension increases the minimum level of rigor to a different extent in each market. In the case of the suspension of the two and three globally most lenient certifiers, there are even three markets in which those certifiers are predicted to be more rigorous than remaining once, such that the minimum levels of rigor in those markets do not change.

parameters remain constant.[60]

Third, I compute relevant statistics for the baseline and the counterfactuals in 2019.

1. The sum of FMUs' weighted averages of expected violation reports, henceforth referred to as the *unconditional weighted sum*, is a proxy of the aggregate quality changes induced by FSC among all once-certified FMUs. Each FMU's weighted average is the sum of the violation reports the FMU expects from each certifier, weighted by the probability of choosing that certifier. The number of violation reports is zero if the FMU does not participate in certification. The sum of those weighted averages is, thus, $vsum \equiv \sum_{it} \sum_{j \in J_t} \hat{s}_{ijt} \widehat{ExpectedRigor}_{jt} \widehat{E[v_{i1t}^0 | \boldsymbol{f_{it}}]}$.
    i. The *conditional* weighted sum $vsum_{FSC}$ is obtained by replacing $\hat{s}_{ijt}$ with $\hat{s}_{ijt|FSC}$, thus, weighting by the probability of choosing each certifier conditional on participating in FSC certification. It is, hence, a proxy of the quality changes induced among certified FMUs, i.e., the quality of certification.
    ii. The weighted sum of expected violations that remain *unreported* due to reductions in participation is the difference between the unconditional and conditional weighted sums.
2. Certifiers' surplus from FSC certification is defined in model (8).
3. FMUs' surplus is FMUs' maximum utility in dollar. For the nested logit model (5), this is $\frac{1}{\alpha_{it}} \left\{ \ln \left[1 + \left(\sum_{j \in J_t} \exp(V_{ijt}/\lambda) \right)^{\lambda} \right] + C \right\}$ (McFadden, 1977). The unknown constant C is the utility from the outside option, forestry without FSC certification. The remaining part, thus, constitutes FMUs' surplus from FSC certification.
4. A measure of consumer valuation of the counterfactual changes is obtained as follows.
    i. I derive a proxy for consumer valuation of FSC certification per certified hectare at baseline, denoted as CV/ha. To do so, I employ the median stated willingness to pay 10% more for FSC-certified products (Cai and Aguilar, 2013). Assuming that this median value has not decreased until 2017 allows using estimates of wood production per certified hectare, which exist only for 2017. Given those estimates and international wood prices, the estimated consumer valuation per hectare 2017 is approximately 15 USD.[61]

[60] In particular, I do not consider potential effects on rigor above the minimum level, FMUs' true violations and the benefits of certification for FMUs, since the data do not allow me to identify these.

[61] Since the surveys summarized by Cai and Aguilar (2013) ask consumers about their willingness to pay a premium for certified final goods, I also need to assume that the valuation of certification for wood products is proportional to the valuation of certification for rough wood. For wood prices, I use a weighted average of 2017 prices per cubic meter of rough wood from UNECE countries (UNECE and FAO, 2023). I compute the average production of cubic meters per hectare from FSC estimates for 2017 (FSC, 2018b).

ii. From that, I deduce $CV/ha/vsum_{FSC}$, the valuation per hectare per unit of certification quality, proxied by the conditional weighted sum of expected violation reports, $vsum_{FSC}$, 0.02 USD in 2017. I assume that this baseline valuation and the production per hectare remain constant until 2019 and across counterfactuals.

iii. I assume that consumer valuation of FSC certification in the counterfactual is the product of the baseline valuation $CV/ha/vsum_{FSC}$, the changed quality of certification in the counterfactual, $vsum'_{FSC}$, and the changed number of certified hectares, ha': $CV' = (CV/ha/vsum_{FSC}) * vsum'_{FSC} * ha'$. That is, consumer valuation is perfectly elastic to changes in quality.

5. I compute welfare changes as the sum of changes in FMUs' and certifiers' surplus and the changes in consumer valuation. I use consumer valuation as a proxy for social benefits. An underlying assumption is that price premia do not change, such that consumer valuation of the changes translates 1:1 to increases in consumers' surplus, instead of increases in FMUs' surplus.

I discuss the plausibility and implications of all relevant assumptions after presenting the results.

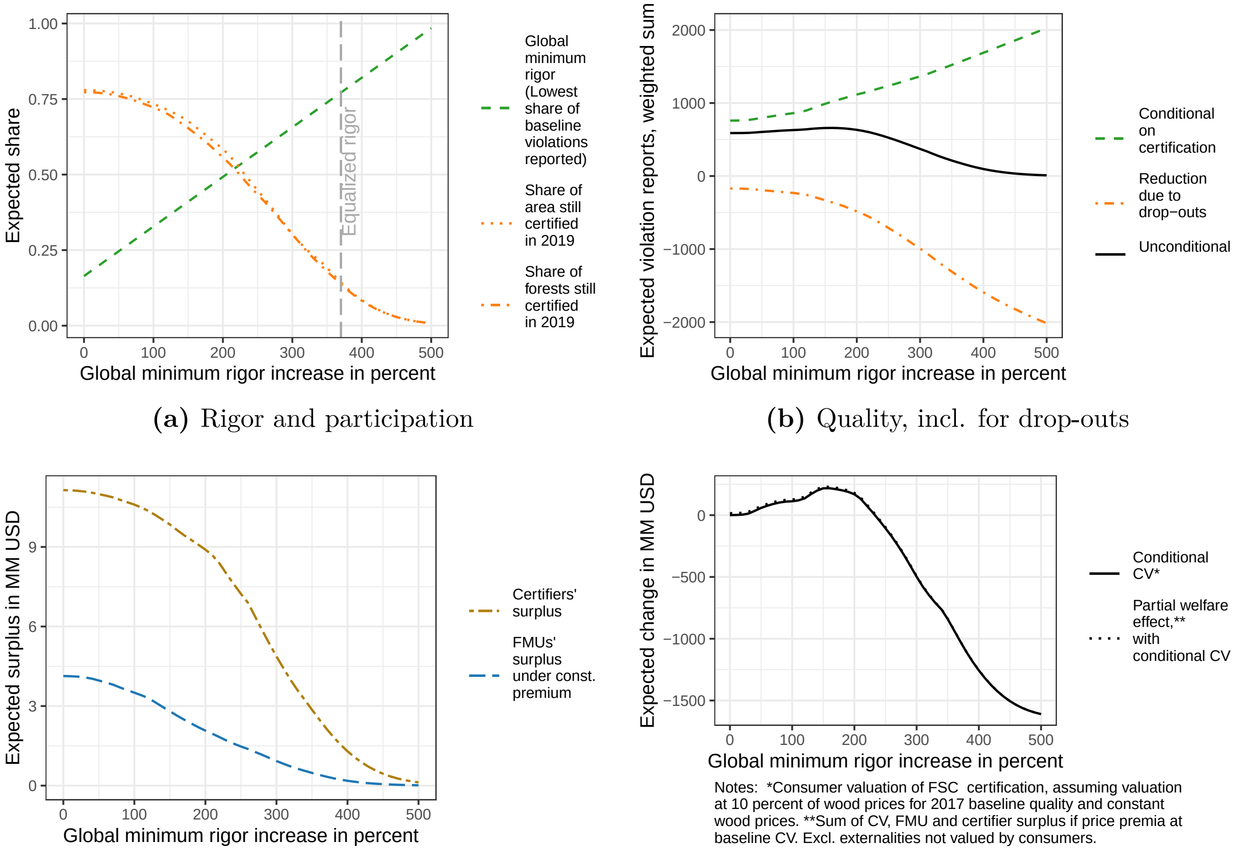
6.2 Results and discussion

6.2.1 Shifting the minimum level of rigor

Figures 4 show how direct shifts of the global minimum rigor and consequent changes in FMUs' certifier choices and prices can affect market outcomes, everything else equal. In each figure, the x-axis represents the enhancement of global minimum rigor in percent of the global minimum rigor at baseline. The global minimum rigor at zero enhancement is the baseline value, the estimate used in the model thus far.

Figure 4 (a) shows the trade-off between rigor and participation. The dashed green line represents the new global minimum rigor in each counterfactual. It rises by assumption. The orange lines show two measures of participation. The dot-dashed orange line plots the share of once-certified FMUs participating in 2019. The dotted orange line plots the share in terms of the certified area. Both develop very similarly. At baseline, 23% of once-certified FMUs no longer participated in 2019. Doubling global minimum rigor leads to only 4 percentage points more expected drop-outs. However, participation drops steadily at an increasing rate for moderate to large enhancements in minimum rigor. When minimum

Figure 4: Expected changes following direct shifts in global minimum rigor



(a) Rigor and participation

(b) Quality, incl. for drop-outs

(c) Surplus from FSC certification

Notes: *Consumer valuation of FSC certification, assuming valuation at 10 percent of wood prices for 2017 baseline quality and constant wood prices. **Sum of CV, FMU and certifier surplus if price premia at baseline CV. Excl. externalities not valued by consumers.

(d) Parametrized change in consumer valuation (CV) rigor more than triples, participation drops very fast at a decreasing rate. Participation approaches zero when the new global minimum exceeds the maximum rigor at baseline. The long-dashed grey vertical line shows this threshold, to the right of which expected relative rigor is equal across all certifiers and markets.

Figure 4 (b) shows the predicted effects of the rigor-participation trade-off on quality in the industry. It translates the changes in Figure (a) to those in the weighted sums of expected violation reports. I interpret these numbers as proxies for compliance with the FSC standard and, thus, quality, particularly for reducing negative externalities of wood production. This interpretation is appropriate as long as most reported violations are corrected and wood production does not change significantly apart from these corrections. The dashed green line shows the weighted sum conditional on participation, a proxy for the quality of certification. It rises at an increasing rate with growing minimum rigor. Doubling the global minimum implies an increase of 13%, equalizing rigor across certifiers an increase of 167%. The orange dot-dashed line shows the weighted sum of expected unreported violations due to decreasing participation. This number drops more slowly than the quality of certification increases, until minimum rigor increases by 160%. At this point, the unconditional weighted sum of expected violation reports reaches its maximum, as the solid black line shows. Beyond this, the unconditional effect decreases and becomes negative for extreme changes in global minimum rigor. But at the optimum, the simulation suggests 70 additional expected violation reports, an increase of 12%. This suggests substantial scope for improving quality in the industry by increasing minimum rigor, even if consumers' willingness to pay remains constant.

However, Figure 4 (c) shows that improving quality is costly. The dashed blue and the dot-dashed orange lines show FMUs' and certifiers' surplus, respectively. They are similar in shape as participation in Figure (a). My estimated model suggests that certifiers' surplus from FSC forest management certification is approximately 11 million USD at baseline and decreases by 5% if global minimum rigor doubles. If rigor is equalized across certifiers, the profit loss is 81%, according to the model's predictions. Reductions in FMUs' surplus are smaller in absolute numbers but larger in the percentage of their surplus at baseline.

Can the improvements in quality for moderate increases of global minimum rigor justify these costs? The dotted line in Figure 4 (d) plots consumer valuation of the counterfactual changes, following the assumptions outlined above, in Section 6.1. The solid line shows the sum of this valuation and the changes in FMUs' and certifiers' surplus. It suggests that consumer valuation of quality improvements alone may generate positive welfare effects of increasing global minimum rigor, even for substantial increases. The predicted changes in consumer valuation are so much larger than the effects on FMUs' and certifiers' surplus that the dotted and solid lines almost overlap.

Discussion: The exercise assumes that price premia for FSC certification remain constant. If price premia increased, they would compensate for part of FMUs' loss and reduce drop-outs. That reduction in drop-outs, in turn, would reduce certifiers' losses and increase welfare further. However, zero or little changes in price premia are a realistic scenario, particularly in the short run. Less than 50% of consumers in 33 countries have ever noticed the FSC label (FSC and IPSOS, 2023). A lot fewer consumers will have information about the quality of FSC certification and even fewer will perceive any improvements in this quality. My measure of changes in consumer valuation is not supposed to represent changes in perceived utility. Rather, the idea is as follows: the median consumer valuation at baseline is a proxy of consumers' true utility given the quality of certification at baseline and this true utility changes with the changing quality of certification, even if these changes are not perceived.

The plotted welfare effects must, nevertheless, be interpreted with caution. First, the predictions capture the effects in the regions that this paper focuses on. Second, I assume that consumer valuation is a linear function of the amount of certified wood and the weighted sum of violation reports conditional on certification. In Appendix D.1, I discuss this assumption and consider an alternative measure of consumer valuation, generating similar qualitative results.

Moreover, I lack an adequate measure of the statistical uncertainty of those predictions. Normal-approximation confidence intervals (Hansen, 2021) computed from predictions from a preliminary set of 550 bootstrap samples suggest that the welfare effects might be statistically insignificant to a large extent. The predictions from the bootstrap samples also suggest that the effects on aggregate quality for very large rigor changes might be statistically insignificant. However, the scope for improvement through moderate changes and the clear trade-off between certification quality and participation are robust to resampling. Given that the original sample represents the whole population, it is unclear whether bootstrapped confidence intervals adequately measure the statistical uncertainty. For that reason, I do not present them in further detail.

Another point of caution is that the social benefit estimates rely on consumers' stated valuation of FSC certification which may not account for all of FSC's positive externalities. Accounting for externalities that consumers do not value may amplify the predicted changes in welfare in absolute terms. Appendix D.2 derives a rough, partial measures of the welfare benefit of reduced tree cover loss through FSC certification, suggesting similar benefits from violation reporting as my computation of consumer valuation. But a lack of research prevents me from deriving a comprehensive measure of the overall benefits of violation reporting.[62]

In any case, label owners cannot fully monetize and redistribute the welfare benefits to incentivize certifiers to increase rigor through monetary compensation as long as there is no fully functioning market for the externalities. Therefore, label owners and their accreditation bodies threaten to suspend lenient certifiers' market access to increase minimum rigor. The following section investigates the effects of such accreditation suspensions.

6.2.2 Suspending the accreditation of lenient certifiers

This section investigates the effects of an indirect shift in certifiers' expected relative rigor through the global suspension of the most lenient certifiers. For each set of suspended certifiers, I also consider an equivalent, direct shift of those certifiers' rigor to the next most lenient certifier's rigor *in each market*.

Table 13: Simulated changes in audit quality and participation

Counterfactual	Avg. minimum rigor across markets (mechanical change)		Participating FMUs (expected change)	
	in numbers	in percent	in numbers	in percent
(1) Targeting the most lenient certifier				
Accreditation suspension	0.04	9.73	-57.16	-5.92
Equivalent minimum rigor shift	0.04	9.73	-11.13	-1.15
(2) Targeting the two most lenient certifiers				
Accreditation suspension	0.08	20.51	-124.12	-12.85
Equivalent minimum rigor shift	0.08	20.51	-57.21	-5.92
(3) Targeting the three most lenient certifiers				
Accreditation suspension	0.11	29.25	-145.86	-15.10
Equivalent minimum rigor shift	0.11	29.25	-103.32	-10.70

Notes: Changes in counterfactual scenarios compared to the baseline. Accreditation suspension is implemented by removing the corresponding certifier from the choiceset. A direct shift of minimum rigor is implemented by shifting the targeted certifiers' expected relative rigor to the next most lenient certifier's rigor in each market.

Table 13 shows the effects on rigor and participation. The first two columns present the induced changes in the average minimum rigor level across markets, in numbers and percent. Each additional certifier suspension induces an average increase in market-level minimum

[62]The benefit from reduced tree cover loss does not account for benefits from more frequent types of violation reports, such as the protection of workers and biodiversity. There is neither sufficient research about FSC's effect on such protection nor are the credible estimates of the social cost of biodiversity loss.

rigor of roughly 10%. The last two columns show the predicted effects on participation. As in the first set of counterfactuals, there is a clear trade-off between rigor and participation.

The reduction in participation is larger with a suspension than it would be if FSC could shift a certifier's rigor without a suspension: I predict that the suspension of a certifier causes 1.4 to 5 times as many drop-outs as equivalent minimum rigor shifts. This suggests that switching cost and other certifier differences are jointly even more important for FMUs than certifiers' leniency.

Table 14: Expected changes in violation reports

Counterfactual	Conditional on certification	Due to drop-outs	Total	
	in numbers	in numbers	in numbers	in percent
(1) Targeting the most lenient certifier				
Accreditation suspension	62.36	-47.58	14.77	2.51
Equivalent minimum rigor shift	18.62	-14.34	4.28	0.73
(2) Targeting the two most lenient certifiers				
Accreditation suspension	120.59	-98.68	21.91	3.72
Equivalent minimum rigor shift	64.65	-45.26	19.39	3.29
(3) Targeting the three most lenient certifiers				
Accreditation suspension	145.39	-116.74	28.65	4.87
Equivalent minimum rigor shift	133.10	-86.56	46.54	7.90

Notes: Changes in counterfactual scenarios compared to the baseline. Accreditation suspension is implemented by removing the corresponding certifier from the choiceset. A direct shift of minimum rigor is implemented by shifting the targeted certifiers' expected relative rigor to the next most lenient certifier's rigor in each market. These numbers are computed under the assumption that consumers value the effect on violation reports conditional on certification.

Table 14 presents the predicted changes in the weighted sums of expected violation reports, proxies for quality. The first column shows the changes conditional on participation, and the second shows the reduction due to drop-outs. The latter follows directly from the reduction in participation, discussed in the last paragraph. The last two columns show the total effect in numbers and percent. The total effect is the effect on the unconditional weighted sum of expected violation reports, a proxy for quality among all once-certified FMUs. I predict positive total effects on such quality, even when all three of the most lenient certifiers are suspended.

In all three sets of suspensions, quality conditional on participation increases more than predicted by an equivalent minimum rigor shift. FMUs' preferences for certifier characteristics other than rigor and switching cost can explain that. Removing a certifier from the choice set forces FMUs to transfer to another certifier or drop out. Some FMUs then transfer to a certifier even more rigorous than the next most lenient due to preferences for characteristics other than rigor. In the cases of suspending the most or the two most lenient certifiers, the effect is more important than the amplified reduction in participation. The total effects on quality among all once-certified FMUs due to suspension exceed the quality effect of an equivalent minimum rigor shift. These results suggest that suspensions are not always a necessary evil compared to direct rigor shifts, but the former can sometimes outperform the latter.

Table 15: Expected change in surplus from FSC certification

Counterfactual	FMUs in MM USD	Targeted certifiers in MM USD	Untargeted certifiers in MM USD	Consumers in MM USD	Total in MM USD
(1) Targeting the most lenient certifier					
Accreditation suspension	-0.51	-1.53	1.20	23.91	23.06
Equivalent minimum rigor shift	-0.10	-0.21	0.11	16.01	15.81
(2) Targeting the two most lenient certifiers					
Accreditation suspension	-0.92	-2.82	1.88	16.42	14.56
Equivalent minimum rigor shift	-0.47	-1.08	0.47	26.94	25.86
(3) Targeting the three most lenient certifiers					
Accreditation suspension	-1.07	-3.48	2.36	7.96	5.77
Equivalent minimum rigor shift	-0.87	-2.38	1.46	49.14	47.34

Notes: Changes in counterfactual scenarios compared to the baseline. Accreditation suspension is implemented by removing the corresponding certifier from the choiceset. A direct shift of minimum rigor is implemented by shifting the targeted certifiers' expected relative rigor to the next most lenient certifier's rigor in each market.

Finally, Table 15 presents the predicted changes in FMUs' and certifiers' surplus and consumer valuation. Unsurprisingly, FMUs and targeted certifiers lose, while the other certifiers benefit from the simulated changes. Consumer valuation greatly exceeds these costs. The predicted effects on welfare are, therefore, positive in all cases. The limitations discussed in the previous section apply here as well. Under the assumptions made, the welfare benefits of suspending the most lenient certifiers may be in the tens of millions of US dollars. Throughout the counterfactuals, FMUs and certifiers lose more through suspensions than through equivalent shifts in minimum rigor, even though untargeted certifiers win more through suspensions. Consumer valuation and total welfare effects are larger for the suspensions of the most lenient certifier than for an equivalent rigor shift, but smaller in case of additional suspensions.

In Appendix D.3, I show the results from equivalent counterfactual exercises that differ only in terms of keeping prices fixed. The results are similar to those discussed above. Price changes due to changes in market power are, thus, only a minor driver of the abovementioned effects. However, price changes reduce the positive quality and welfare effects of suspensions and, albeit to a much smaller extent, of equivalent rigor shifts.

Overall, the presented counterfactuals show that stricter accreditation has the potential to reduce negative production externalities and increase quality and welfare. Reduced participation decreases the positive effect. Suspensions reduce participation, FMUs' and certifiers' surplus more than do direct shifts in minimum rigor, if they were feasible. However, suspensions also have positive side effects and lead to greater welfare benefits in some cases.

7 Conclusion

Certification is used to check compliance with quality standards whose implementation is costly and not observed by consumers and investors. If firms maximize profits, basic theory would not predict them to comply more than they need to obtain certification. When all certifiers of a given standard, such as FSC, provide the same signal to the public, firms are expected to prefer more lenient certifiers since leniency reduces compliance costs. The results of this paper confirm this but also show that other factors may mitigate the mechanism. In the context of this paper, for example, a forest unit's certifier choice is very persistent due to high switching costs. That reduces the degree of shopping for leniency. Nevertheless, this paper suggests that firms' willingness to pay for leniency provides bad incentives for certifiers that reduce the credibility of certification, unless counterbalanced by strict accreditation.

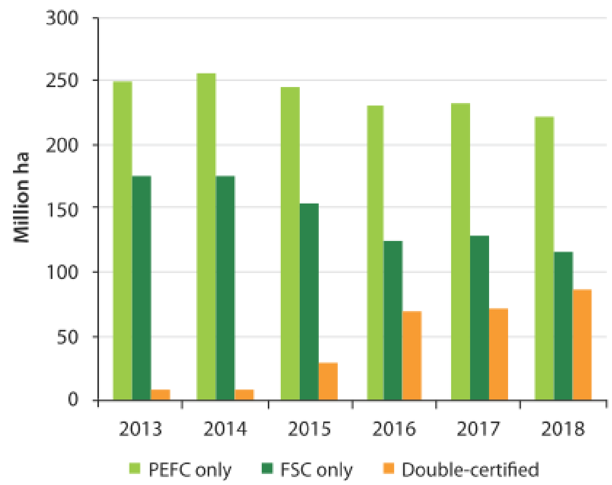
This phenomenon has not only led to calls for stricter accreditation, but also for reorganizing certification, for example, by randomly assigning certifiers and paying them through a fund (Earthsight, 2021). Yet, the reason FSC is hesitant to do so might be the risk of increasing forest units' cost beyond their participation constraint. The framework and data of this paper are insufficient to evaluate the effects substantial structural changes such as financing certifiers through a fund would have. But my counterfactual analysis shows that stricter accreditation by suspension of lenient certifiers would indeed reduce participation and the surplus of forest firms and certifiers, as long as price premia do not change. Nevertheless, I predict that the overall effect of moderately stricter accreditation on quality and welfare would be positive.

FSC and any institution regulating voluntary certification face the trade-off between participation and quality of certification. Empirical estimates of demand parameters and measures of leniency are crucial to evaluate the scope for improvement within each context. This paper contributes to that by providing a framework that may be applied in other contexts, where audit results are available. The paper has limitations in the identification due to multiple selection issues and limited price data. A helpful extension would be to model and estimate the drivers of certifiers' choice of rigor in a dynamic context of suspension risk. Such an extension would require a longer time horizon and richer data to estimate region-level rigor. In addition, data on price premia and purchases of certified products by upstream firms and consumers could allow to estimate how benefits of FSC certification for FMUs vary with changes in the quality of certification and demand-side parameters.

Understanding the trade-off between participation and quality is also crucial for policy makers that want to take complementarities of voluntary certification and mandatory regulation into account. This paper shows that while there is scope for improvement in case of FSC, the certification scheme could most likely not afford very large increases in audit rigor. To reduce negative production externalities beyond those limits, price premia would need to increase substantially or corresponding rules would need to be mandatory.

A Appendix: Institutional setting

Figure A1: FSC and the PEFC certified forests areas, 2013-2018 (UNECE/FAO 2019)



Note: Figure A1 shows that an increasing share of forests is certified both according to Programme for the Endorsement of Forest Certification (PEFC) and FSC standards. This phenomenon may be explained by the demand from different downstream firms that buy from those forests and may have focused their communication on one of the two standards. That only affects the analysis if forest units choose their certifier jointly for FSC and PEFC. But even then it should not affect the main results, as certifiers are probably of a similar rigor type in PEFC and FSC certification.

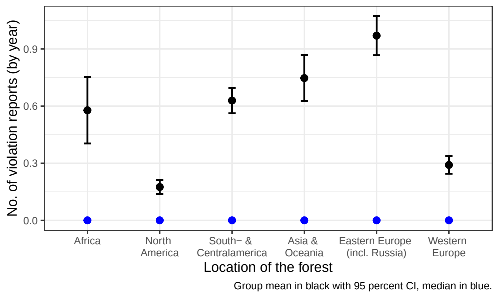
B Appendix: Data and descriptive evidence

B.1 FMU panel

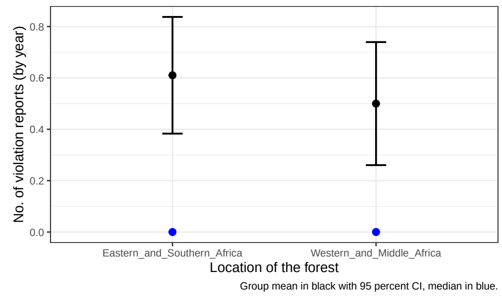
B.1.1 Audit data extraction

In this section, I describe the extraction of the full audit data collected for the FMU panel, i.e. without restriction to the regions that the analysis focuses on. The full dataset contains basic information on the audits conducted in 2015-2020 with FMUs whose Forest Management / Chain-of-Custody certificate was valid at some point between 2015 and 2020, to the extent that they are documented in the audit summaries published on the certificate profiles

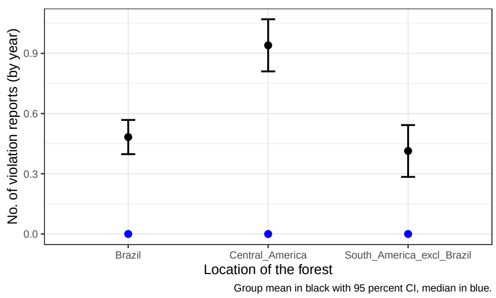
Figure A2: Major violations reports in FSC audits 2015-2019 across regions



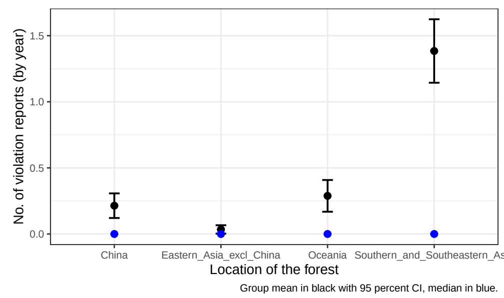
(a) Global comparison



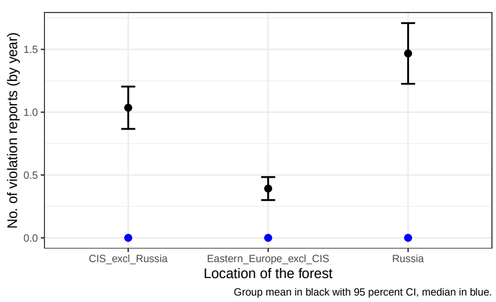
(b) Africa



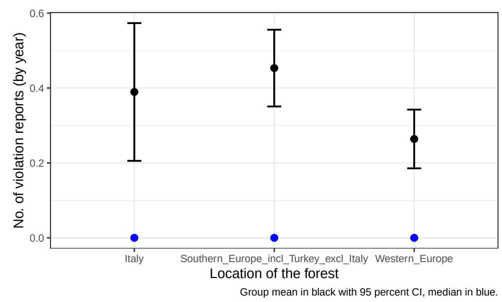
(c) Latin America



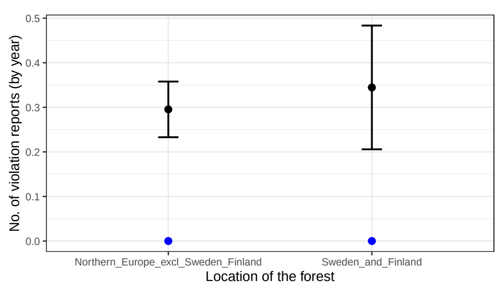
(d) Asia



(e) Europe (1)



(f) Europe (2)



(g) Europe (3)

Figure A3: Certificate list in FSC's public database (FSC 2023)

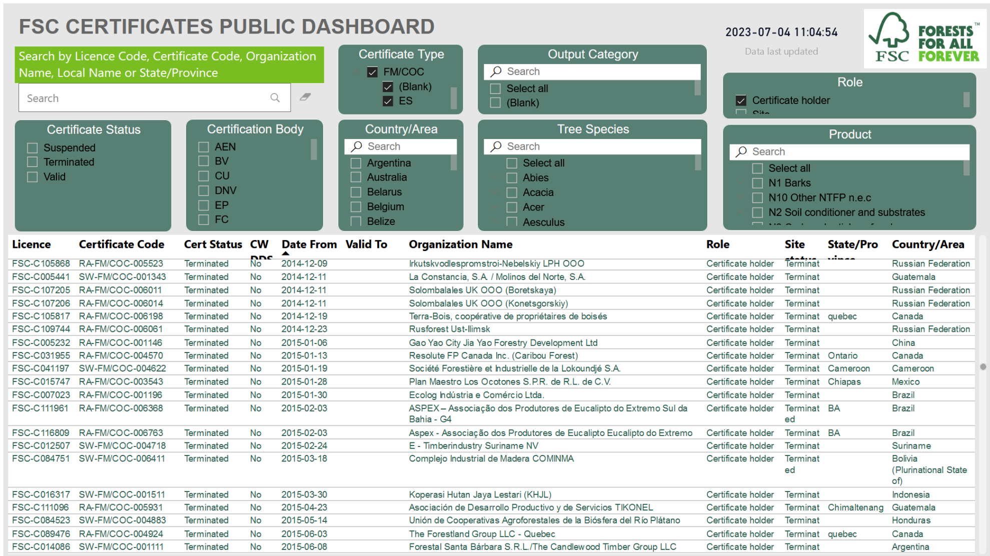
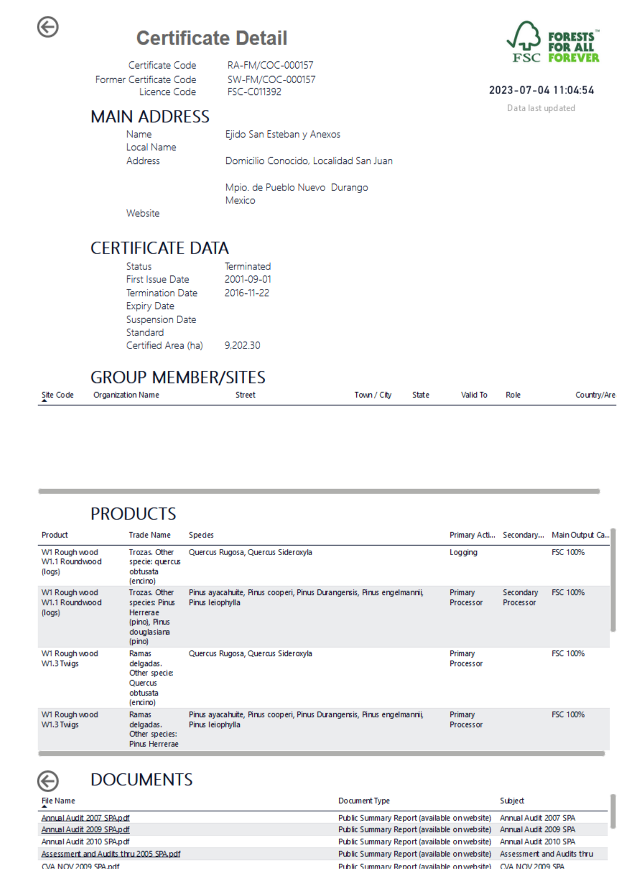


Figure A4: Certificate profile in FSC's public database (FSC 2023)



accessible through https://info.fsc.org/certificate.php, as demanded in FSC-STD-20-007b (V1-0) EN.

Scraping of the raw data

From the FSC Certificate Search, I scraped basic information about all FMUs and the links to their profiles. Basic information on the certificates and the files published on the profiles were downloaded first in June 2020[63] and then in September - November 2021.[64]

Filtering one report per audit

From all the files, the final dataset should only contain one summary for each audit. In case of several files referring to the same audit, the filtering process aimed to choose (1) the "oldest" or most original summary, i.e., the summary with the report date that is closest to the audit, and (2) the summary in the most common language from the perspective of the Researcher.

The motivation for choice (1) was that there were cases where corrected violations were no longer mentioned in later versions of an audit summary. The motivation for choice (2) was first to facilitate the correct extraction of variables by preferring languages that were either known to Researcher (English, French, German) or which were more prevalent (Spanish preferred over Portuguese, Russian over Ukrainian).

The files were filtered with algorithms built according to those rules by the following means:

1. Files with no audit summaries ("Public Summary Report"), according to the information on the certificate profile or clear indicators in the file name, were removed, such as member lists and species lists.
2. For other files, different versions in time or languages could be identified by certain suffixes used in the file names by certain certifiers, e.g., "_ENG.pdf" for English or "_V2.pdf" for a second version in time of the same report.
3. The language of the file was identified using language-detecting algorithms from the R packages *cld*2 and *cld*3. For two identical or almost identical file names, with one file being in English and the other in a local language, the one in the local language could

[63] The certificate profiles were scraped the 04/06/2020, including the links to the files published on the certificate profiles. From those links, the files were downloaded from 04/06/2021 until 01/07/2021. The long time scope for that was taken to avoid overcharging the FSC portal.

[64] The list of certificates from the FSC Certificate Search, including the certificate profile links, was scraped on 30/08/2021. Information from the certificate profiles, most importantly the file links were scraped on 02/09/2021. Yet, due to an error noticed later, additional information from the certificate profiles, such as the certified forest area, was incompletely scraped, so the data from certificate profiles in this final, complete dataset comes from scraping the profiles the 11/11/2021.

then be removed. Languages are privileged in the following order: English, Spanish, French, German, Russian

4. Once audit and report dates were extracted, these were used to identify different summaries of the same audit. The identified certifier and audit type were also used as further confirmation of identical audits.

Variable extraction

Relevant variables were extracted from the audit summaries using automated extraction with regular expressions and manual addition wherever the former was too complicated or risky. Based on the extraction of the audit year, the dataset was restricted to reports on audits conducted from 2015 to 2020.[65] The information extracted from one audit summary concerns on the most recent audit covered in a report (see the discussion in the section on Completeness). The following subsections document choices made in the extraction.

The certifiers are identified from the certificate code in the document. If this is not found (majority of the files), it is identified from mentions of the certifier in the full name or sufficiently unambiguous abbreviation on the first page. Suppose this is not found (a few hundred cases, in particular, where the certifier only appears in a label in image format). In that case, the certifier is identified from patterns in the naming of the file or manually added.

Audits are classified as (Re)certification audits (Initial or Recertification), Annual surveillance audits, Audits to verify the correction of major violations, Other special audits. Audit types (3) and (4) are excluded in the analysis. The types are mostly identified from the title or corresponding table on the front page of the file. In other cases, clear patterns in the file name indicate the type. For a small remaining part, the types were added manually.

The report date that was aimed to extract is the most recent update date of the audit summary, as it should be published on the front pages. Where the algorithm finds no report date, dates from the file names are used if sufficiently clear. The audit years are extracted where possible from the front pages and, if not possible, looked up in the file, typically in the section of the audit schedule.

The violation reports were collected in the following steps. I aimed to include all upgraded major violation reports as new violation reports.

[65] Temporarily, also reports from 2014 were included to remove observations from 2014 that appear again in 2015, as will be explained for the major violation reports below.

1. Where possible, with sufficiently clear identifiers, violation references are extracted into the violation list. The number of such references is counted as the number of violation reports. Upgraded violations that were clearly marked as such obtain the suffix “upgraded” in the violation list.
2. Where this is too complicated or risky but a section in the file states how many violations of each grade were found, the number of violations is extracted directly from there.
3. The rest of the violations are added manually.
4. For violations collected in step 1 and from certifiers where it is not always possible to exclude violations from previous audits, violations from previous audits were removed ex-post in a dedicated algorithm:

- First, by excluding clear violation references that appear in several audit summaries of the same CH are removed from the newer audit summary. Wherever the violation reference is not clear enough to be unique, we check manually whether the reference in the same document refers to the same violation or not (e.g., for CB 2 the format of the corresponding table often only allows extracting simple numbers of violations such as “1”.)
- Second, by excluding those violations whose patterns suggest that they were obtained in a previous year (CB 1(a) for example, typically denotes violations as “01/18” in 2018, which the algorithm would remove for an audit of the year 2019), unless if the reference had the suffix “upgraded”. For CB 1(a), CB 1(b) and Imaflora, where upgraded did not have clear markers and would thus be wrongly removed, all files were violation patterns from the previous year (e.g., “01/18” for an audit in 2019) were rechecked manually to add upgraded violations where necessary. For CB 7, where upgraded violations are not marked but not extracted in the algorithm above, all documents were rechecked, and upgraded violations were added where necessary.

5. Each file with duplicate violation references was checked manually to remove the duplicates in case they refer to the same NC. This may not be the case, either if the algorithm did not extract the complete violation reference[66] or if the certifier uses two identical references for different violations (recognized by having a separate box with a different description and concerned indicator of the NC).
6. In all cases where only the number of violations was extracted but was different from

[66] E.g. if the rest of the reference appears in the next line after text in other columns in a table that is not actually formatted as a table as often for CB 2.

zero, the references were added later when adding the concerned indicator/norm.

For easily identifiable and common combinations of certifier and language, the indicator or norm/reference text field was extracted automatized for each reported major violation, which was detectable with regular expressions. For all others, the text fields were added manually. The text fields also include the standard references for some certifiers but not for others. Yet, given that the numbering of national FSC standards shall follow the international Principles and Criteria, the indicator should allow to identify at least those. This is not true in some cases, potentially due to exceptions to the mentioned rule. Yet, the applicable national standards should be identifiable, given the information on the country and year of the audit and information on the FSC website on standard changes (https://fsc.org/en/document-centre). From the text field, the indicators were extracted using the standard pattern of the principle number being followed by a dot and the criterium number, potentially again followed by a dot and the indicator number (in some cases, the dot is further followed by a "p" for Principle or "c" for Criterium etc.).

B.1.2 Audit data accuracy and completeness

This section provides a quick evaluation of the extent to which the dataset matches the information published on the certificate profiles and in the audit summaries, as far as we have been able to assess it, given time constraints. Further possible assessments are suggested.

Completeness

The datasets of the FSC certificate search and, consequently of the certificate profiles scraped in 2020 and 2021 are complete: After scraping, they contained the same number of observations as found online (checked at the time): 3532 in 2021 and 3309 in 2020. After removing duplicates and certificates that were terminated before 2015, there are 2506 in 2021 and 2277 in 2020.

Possible reasons for the missing audit summaries in the final dataset include the following:

- Audit summary is not or no longer on the certificate profile (In many cases, we could see that after a certifier switch, the reports from the previous certifier are no longer available (even though they should have kept them))
- Unaccessible files (not downloadable / behind a login).
- Corrupted / not machine-readable files

- Filtered by the algorithm due to inaccurate information in the audit summary (typos in the audit dates etc.)
- Filtered by the algorithm due to inaccuracies in the algorithm itself

Given that there are exceptions to the necessity of an annual audit in the FSC standard and that certifiers are allowed to join reports of several annual surveillance audits in the same document, it is not straightforward to check whether audit summaries are missing. In most of the latter cases, a new version of the audit summary is published for each audit, so we always decided to extract only the information on the most recent audit covered in a report. Given this choice, information from former audits may be missed if the certifier removed the former version of the joint file of audit summaries.

Scrape 2020:

The total number of files of the certificate profiles at the time of the download was 22730. 27 files could not be accessed and downloaded, partly due to a wrong link or requiring a login. 158 were identified as duplicates and removed. The remaining files were read into a text format using the R package *readtext*. Excel files were read into as lists of data frames (one per sheet), using the R package *readxl*. For 191 downloaded files, no text could be read. Those files are corrupted.

Scrape 2021:

2021 we only downloaded the files under URLs that were not contained in 2020, 7578 in total. 694 could not be read in or were duplicates.

After removals of non-audit files, newer versions of the same audits and translations, we have in the merged dataset around 9000 audit summaries. There are around 2000 additional CH-year combinations (not included in the dataset) without an audit summary[67]. Yet, in a good part of cases, there are instead two audits in the previous year. The summary seems missing in others, potentially for one of the reasons noted above. Further checks would be necessary to ensure complete audit sequences.

It would be useful, for example, to check for 100 randomly chosen certificates, to what extent, and why audit reports were missing compared to the files published on the certificate profiles. In the current project, there was no time for this though.

Accuracy of extracted data

[67]Excluded in the final dataset as that is organized at the audit summary level.

For 100 randomly chosen files, the accuracy of the extracted variables was checked manually (accuracy according to the information found by looking into the report). For the first set of variables, this was checked on subsamples drawn from the whole dataset, including some fraction of manually extracted.

Variable	Accurately assigned
Certifier	99%
File type	97%
Audit date	99%

For the second set of variables, this was checked on subsamples drawn from only those observations that were not added manually anyway. The approximate fraction of observations of the given variable in the whole dataset, which was added manually, is given in the second column. Assuming all manually added observations are correct, the fourth column computes the overall accuracy according to this check.

Variable	Fraction manually detected	Accurately assigned among obs. extracted with Regex	Total accurately assigned if one assumes 100% accuracy for manuals
Number of major violation reports	approx. 34%	97%	approx. 98%
Number of minor violation reports	approx. 26%	>91%*	>93%
Reported major violation indicators	approx. 36%	>91%*	>94%

*The sources of most mistakes in that test were corrected in the newest dataset version. The accuracy rate is likely much higher now, but there was no time left for another test.

B.1.3 Applicable rules for audit summaries

Public summaries of FSC Forest Management certification audit reports and annual surveillance updates (hereafter "audit summaries") are published according to a set of rules specified in FSC's standard for certifiers and accompanying guidelines (FSC, 2009a, FSC (2021b)).

The most relevant for the data collection and analysis can be summarized as follows:

1. **Content:** Among others, the following variables must be published in each audit summary: The name and contact details of the certifier and the certificate holder, the date the report was last updated, the name and location of the certified forest areas, the date of issue and expiry of the certificate, the dates, duration, and type of the field audit (annual surveillance, (re)certification or correction of violations), the (updated) certification decision and most importantly "a list of all non-conformities [violations] that the managers are required to correct to maintain their certification, including the time course within which corrective actions shall be taken" (See Art. 5, 7, 8 and Box 1 in FSC (2009a)).
2. **Publication date:** The summaries of the certification audits must be published before a certificate is issued / re-issued. Annual updates must be published within 90 days after the surveillance audit. (See Art. 3 in FSC (2009a))
3. **Deletion date:** The audit summaries shall remain in the public database, even after the suspension of a certificate holder (See question INT-STD-20-007b_05 in FSC (2021b)). FSC clarified this in 2020, so there might be inconsistencies before that.
4. **Language:** For certificates that cover more than 1,000 ha forest area and for which not all group member forest units are "small", i.e., less than 100ha each (Or designated as small in a formal procedure outlined Art. 2.2 FSC (2004)) the summaries must be available in one of FSC's official languages, which are. For smaller forests, the local/national language of the forest's location is sufficient (See Art. 2.1 in FSC (2009a))

Certifiers have to upload the audit summaries on a database that is managed by FSC and from which they are automatically published on a public certificate profile that can be accessed through a search engine.[68]

B.1.4 List of yearly country-level characteristics

- The numbers of FSC-affine upstream firms as upstream demand proxies (FSC, 2023b), including certified downstream firms, certified public projects, and retailers officially promoting FSC products.
- Availability of national FSC standards (FSC, 2023a)

[68] The search engine is https://info.fsc.org/certificate.php. An example of a certificate profile accessed through the search engine is https://info.fsc.org/details.php?id=a023300000azGp6AAE&type=certificate

- Trade values and volumes in different wood product categories (FAO, 2023)
- The percentage of publicly and privately owned forests per country in 2015 (FAO, 2020)
- Transparency International's Corruption Perceptions Index (Transparency International, 2021)
- Indicators of progress towards sustainable forest management (United Nations, 2021)
- GDP per capita, profit tax rates, currency conversion factors, and inflation rates (IMF, 2020; World Bank, 2021)
- Geographic characteristics, such as being landlocked, distances, and languages (Mayer and Zignago, 2011)

B.1.5 Variation in violation reports and FMU characteristics

Figure A5: Major violation reports in audits with at least one reported major violation

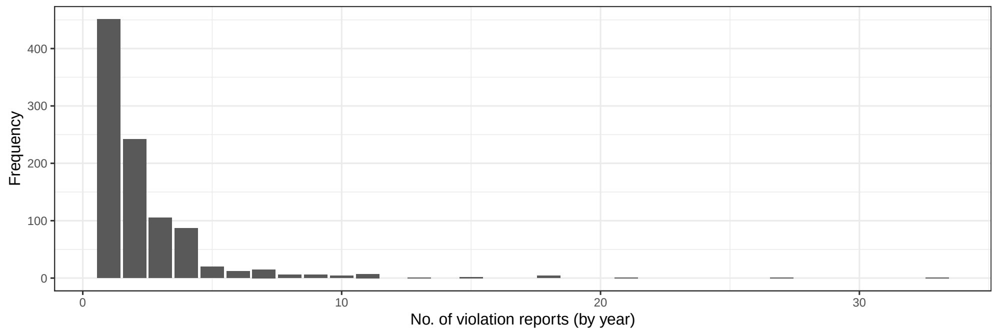
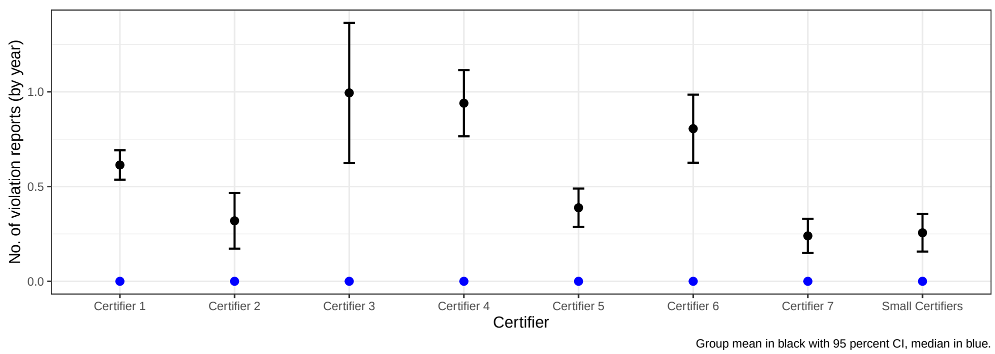


Figure A6: Major violation reports across certifiers



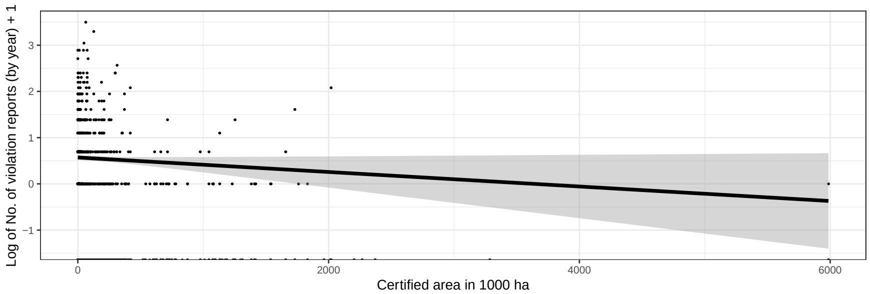
B.2 Market definition, choice sets and certifier-market panel

B.3 Price panel

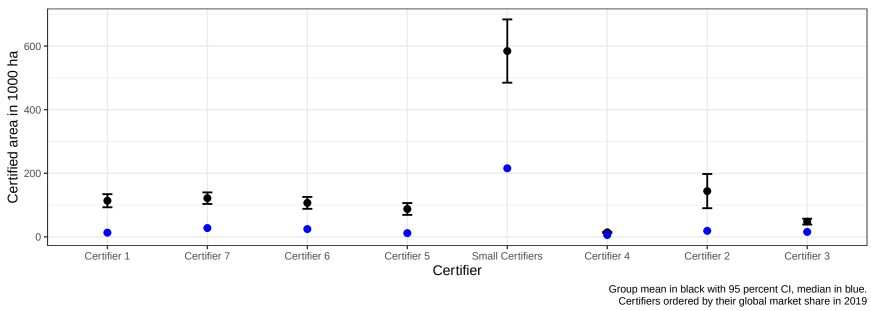
B.3.1 Survey design and response

The survey was sent by e-mail to all FMUs with a valid certificate in June 2020 for whom an e-mail-address could be found. I excluded terminated certified FMUs since they are likely to be less motivated to participate in a survey and might even have an incentive to give false answers out of negative sentiments towards FSC. From the 1756 valid certificates in June 2020, e-mail-addresses of FSC contact persons could be extracted from audit report summaries for 1456 and found through their websites for 109 unique certified FMUs. 101 certified FMUs have two or more certificates (up to 6) for different forest units which may

Figure A7: Selection of FMUs of varying size into certifiers

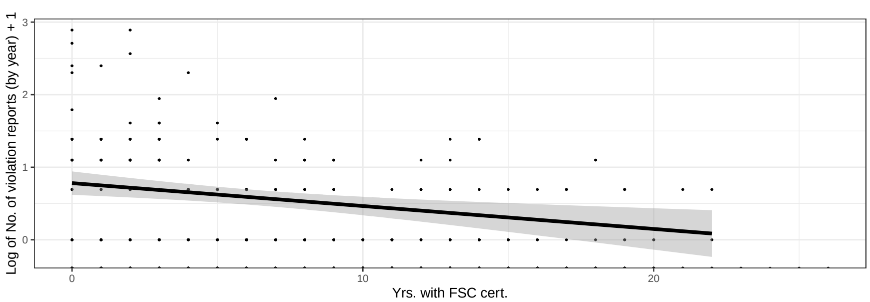


(a) Major violation reports per year in 2015-2019

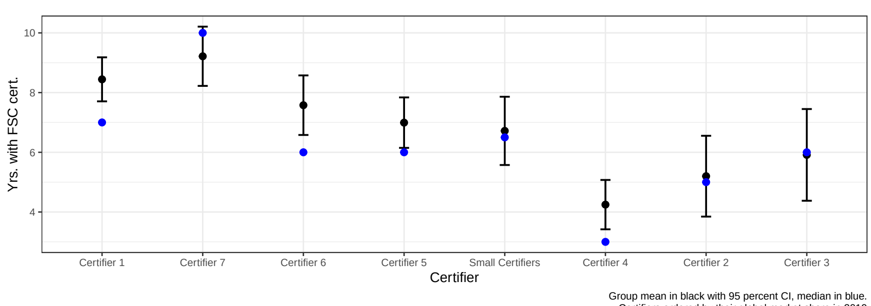


(b) Mean per certifier in 2015-2019

Figure A8: Selection of FMUs with varying certification experience into certifiers

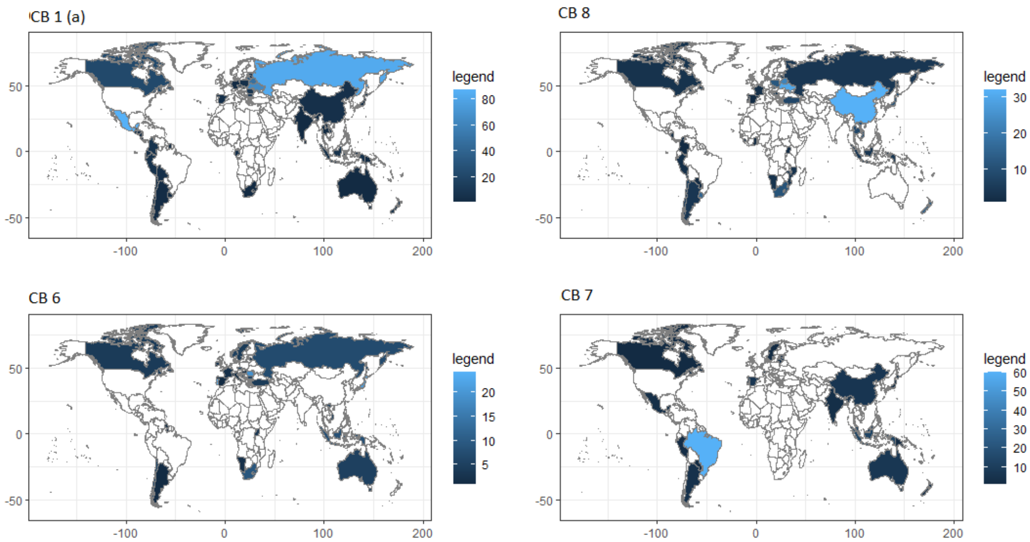


(a) Major violation reports in 2019

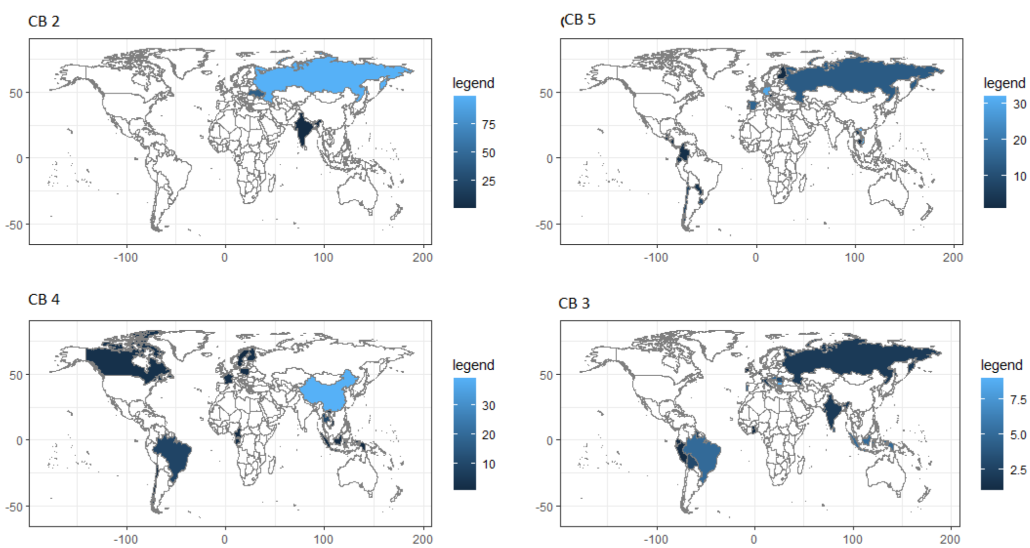


(b) Mean per certifier in 2019

Figure A9: Numbers of valid certificates in May 2020 issued by the largest certifiers (CB) in different countries



(a) Certifiers 1, 8, 6 and 7



(b) Certifiers 2, 5, 4 and 3

be the case if a firm owns forests in very diverse locations. These were asked to respond for each certificate separately but did not follow the request. 12 separate certificates of the same certified FMUs were however added manually ex post based on the information three certified FMUs gave in comments about the multitude of certificates they have / have had in the past. Over 100 e-mail-addresses turned out to be out of date or incorrect, about 40 could be replaced by another contact that was accessible. For 66 cases, no other email address could be found.

The certified FMUs with valid e-mail-adress were invited by e-mail to participate in a short survey asking questions about the prices they had to pay for certification and a few characteristics of them that may be relevant to classify the prices It was conducted in an anonymous manner in order to encourage firms to participate in spite of the sensitivity of the issue.

335 certified FMUs submitted complete responses to the survey, a response rate of 21 Percent.[69] Complete means that they replied at least to all mandatory questions[70] and explicitly submitted their response at the end of the survey. In comparison, response rates in other, much smaller surveys of FSC certified FMUs range from about 15% to about 80% (Araujo et al., 2009; Galati et al., 2017; Jaung et al., 2016; Overdevest and Rickenbach, 2006). These surveys did not face the additional challenge of a multiplicity of languages and locations (not permitting a dual-mode survey) and questions about prices which are particularly sensitive. Moreover, the way in which I use the data from the survey in the main project does not require it to be perfectly representative.

From the remaining respondents,[71] I obtained 1122 informative price quotes or approximations, out of which 877 were prices quoted explicitly.[72] To understand the difference, note

[69] In addition, there were 2 respondents that seem to be not part of the target population: They claimed that their forest had never been certified. This might be due to a mistake in the e-mail-address, an employee that is not aware of the certification or a lack of willingness to respond. They were thus excluded from the survey data.

[70] These can be recognized by a red * before the question, as seen in Figure A10 and the following.

[71] 50 of the respondents had to be removed, 25 of them since an unknown fraction of their certification fee is paid by a donor, while the prices quoted by the other 25 were not informative or reliable for other reasons. 9 of them did not mention their certifier. (7 commented that the price quoted included multiple certificates, either multiple FSC certificates without saying how many or certificates from other standards. 6 did not enter any price quote while 1 entered Zeros everywhere.)

[72] 2 of the included respondents' yearly price quotes were removed because the respondents did not note the certifier for the corresponding year. 11 concerned years in which the certificate holder was not certified and thus paid nothing. 59 price quotes concerned years before an extension of the certificate; these observations were excluded as important characteristics of the certificate such as the certified area were only provided with respect to the period after the latest extension. 4 observations concerned years in which the certificate holder quoted a price for multiple certificates together without specifying the number or mixing with a certificate from another standard. In 6 cases it was noted that they had a certifier which was not in the suggested list

that the survey first asked the respondent to type the price paid for certification in each year (in the following "direct price quotes"). In a second question, respondents were asked to use a slider to mark the approximate price (in 5000 USD intervals or the chosen currency's corresponding interval). Since also the direct price quotes can be approximations, I may use the approximate price quotes in the 245 cases in which no direct price quote was given. In my main specification, I focus however on the sample with the direct price quotes. The approximate price quotes allows for a robustness check.

The price quotes were then transfered into USD prices, using the currency conversion factor used by the World Bank (World Bank, 2020). They were further adjusted across time, using the US paper and wood pulp PPI, with 2015 as base year (FRED, 2020).

To analyze whether the survey data are reliable, I compare the distribution of respondents to the full sample of valid certificates and the price quotes of similar firms to check for consistency in the next sections. I also checked for consistency of denoted world regions, certifier availability in this region, the language in which the survey was conducted and the currencies denoted. Responses were consistent and many respondents raised additional confidence in their motivation to respond thouroughly by additional comments they made in writing.

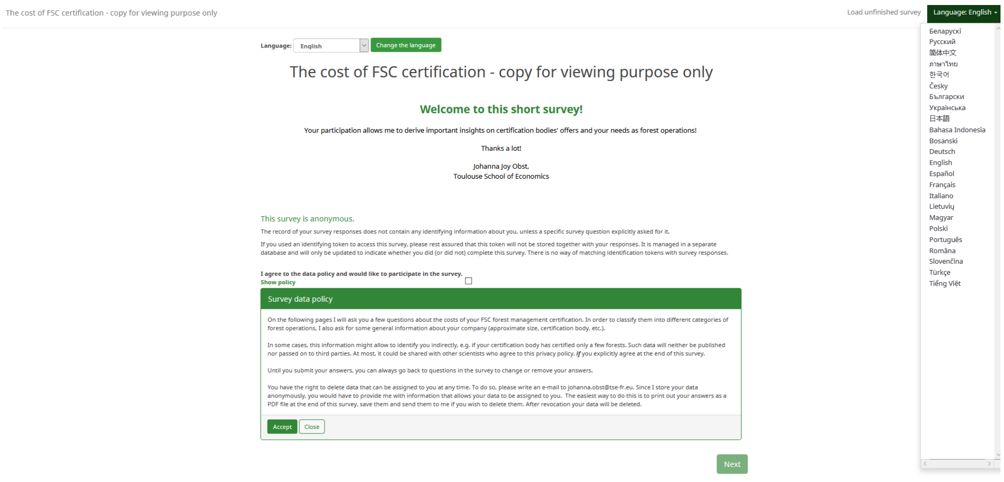
B.3.2 Comparison with the main dataset

A comparison of the distribution of all characteristics that we observe both in the survey and in the certificate dataset suggests that the survey respondents do not differ much from the whole population of certified FMUs. This can be seen in Figures A15, A16, A18 and A17 comparing the distribution of respondents and all certified FMUs across geographic regions, experience with certification as well as size of the forest. Four differences seem nevertheless relevant.

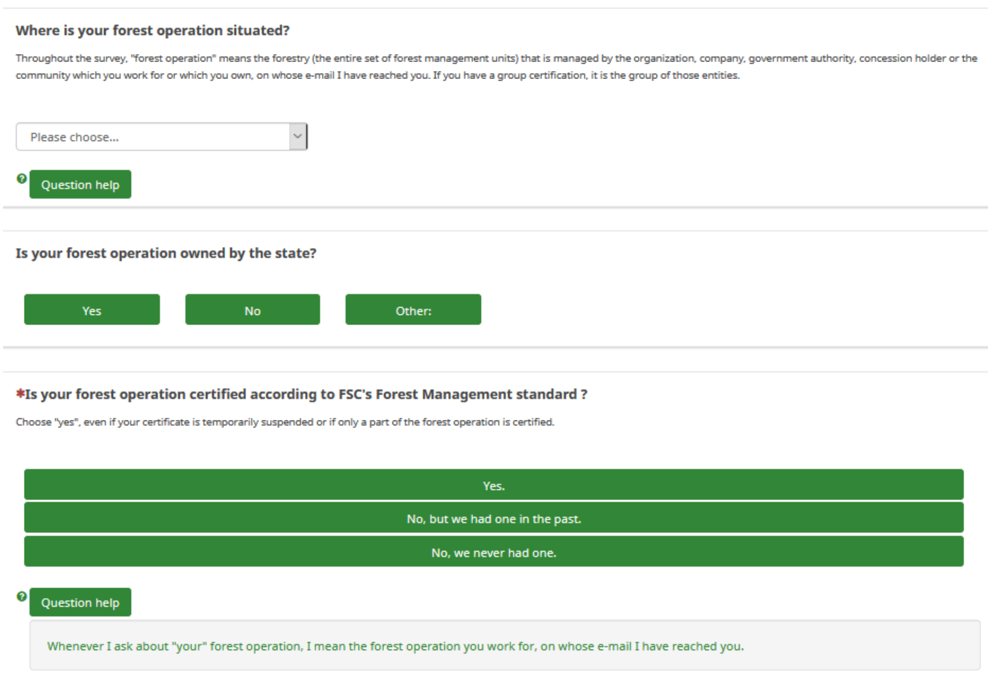
First, Figures A15 show that some certifiers are more representated than others. Especially underrepresentated is CB 8with only 24 informative price quotes. Corresponding estimates will thus need to be interpreted with caution and I thus exclude the countries of the Commonwealth of Independent States in at least one version of my analysis, as it is the only region in which CB 8is present.

without specifying which one. And in 23 cases, a zero price was noted which might not be wrong if no audit was conducted in the year, but my analysis uses only years in which audits are conducted so that such "price quotes" should not be taken into account.

Figure A10: Overview over the survey with valid FSC FM certificate holders (Oct.-Nov. 2020)

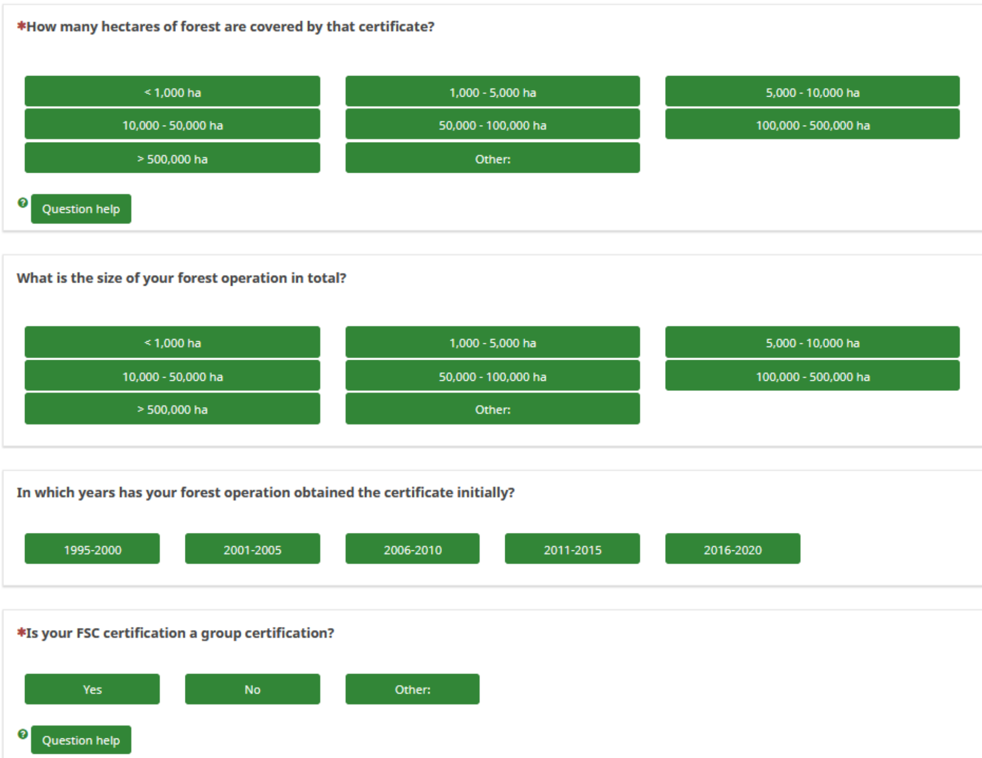


(a)

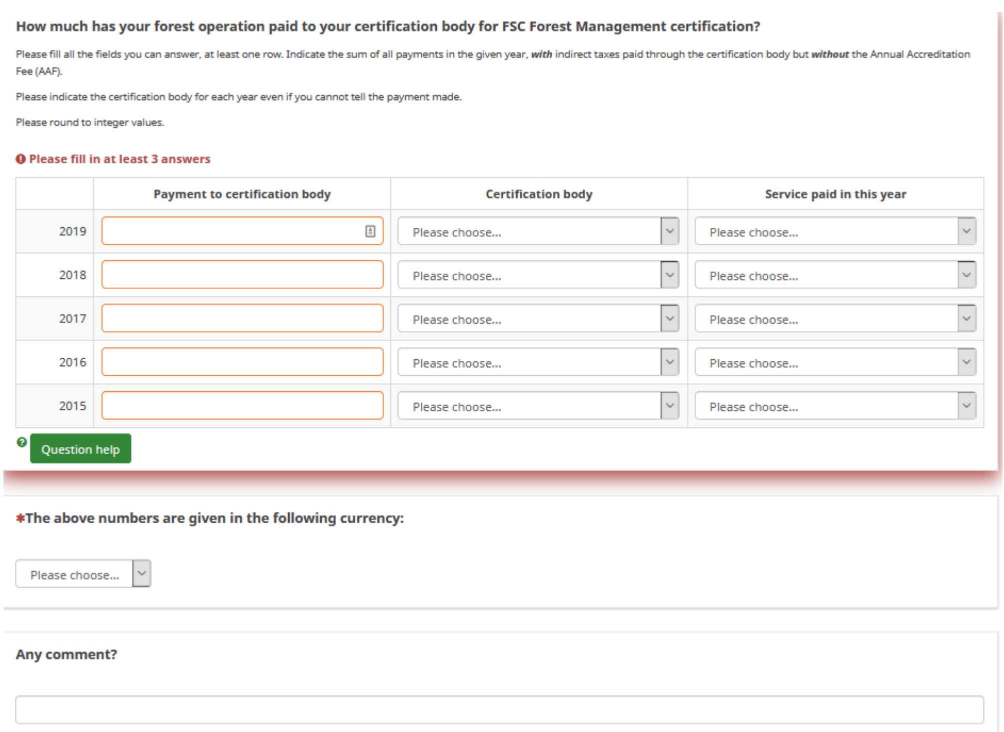


(b)

Figure A11: Overview over the survey with valid FSC FM certificate holders (Oct.-Nov. 2020)

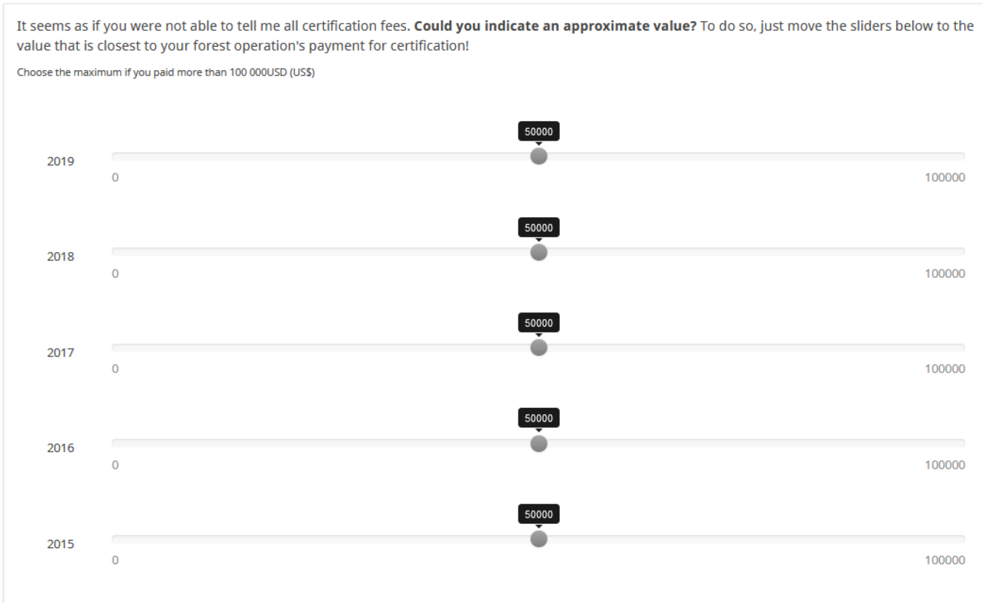


(a)

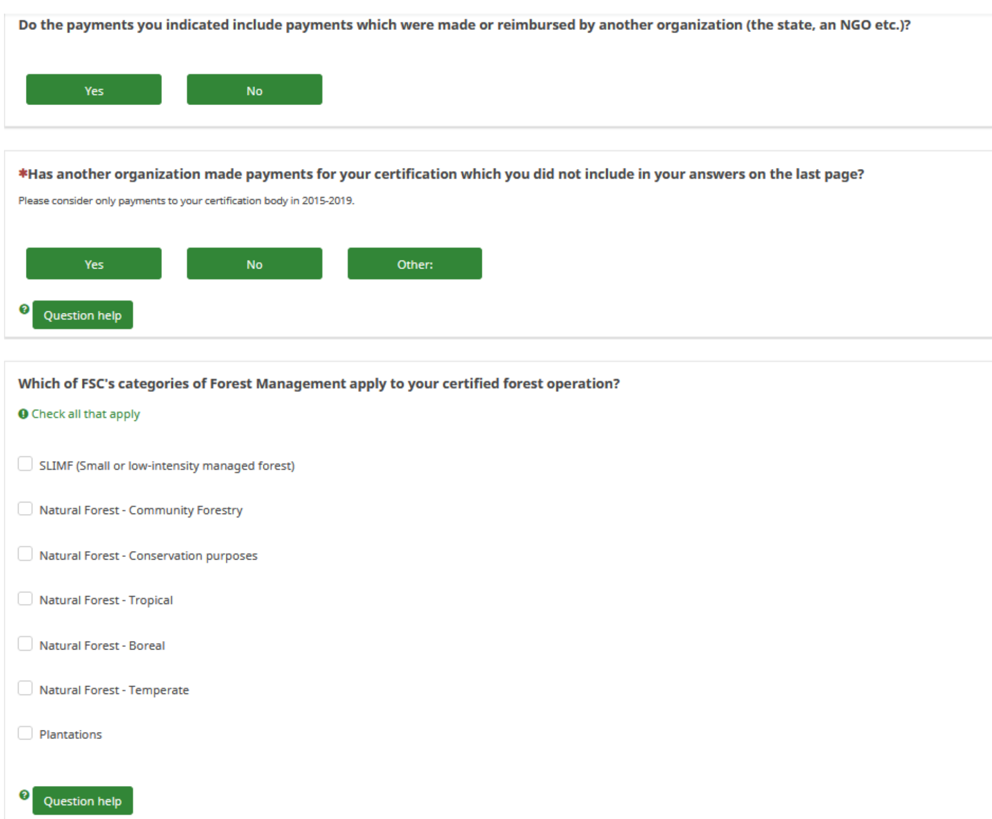


(b)

Figure A12: Overview over the survey with valid FSC FM certificate holders (Oct.-Nov. 2020)

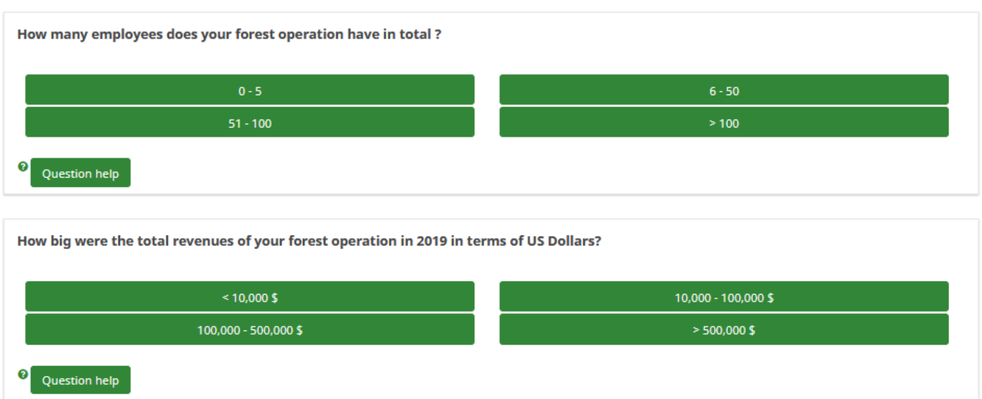


(a)

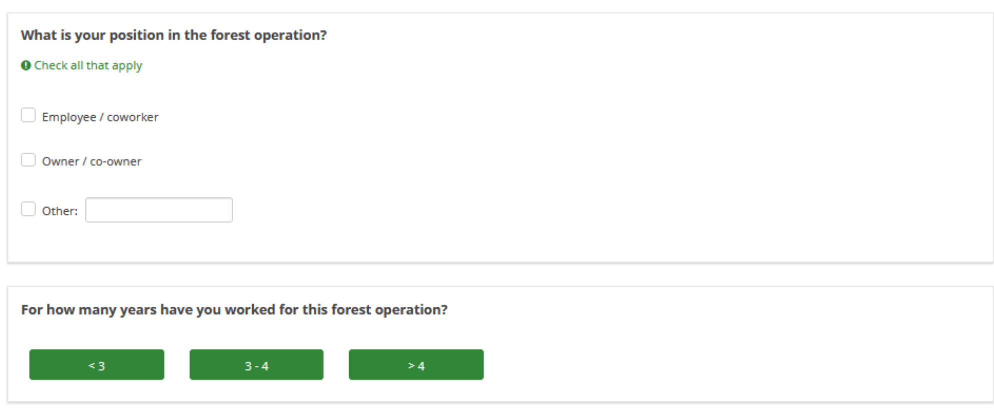


(b)

Figure A13: Overview over the survey with valid FSC FM certificate holders (Oct.-Nov. 2020)

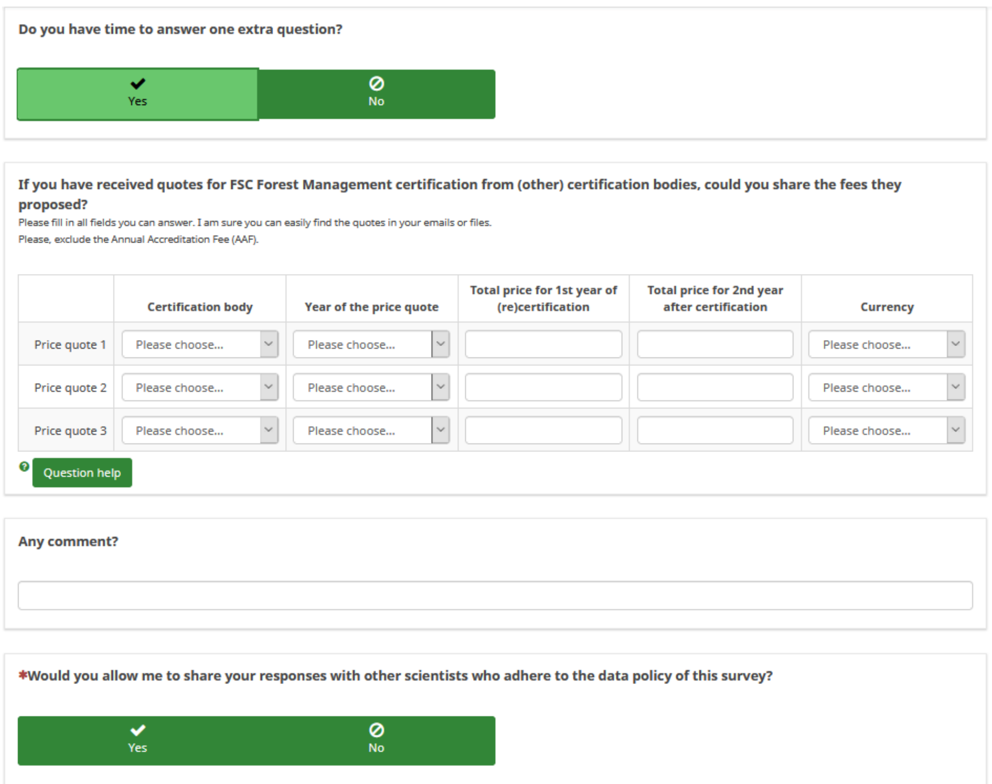


(a)



(b)

Figure A14: Overview over the survey with valid FSC FM certificate holders (Oct.-Nov. 2020)



Do you have time to answer one extra question?

Yes | No

If you have received quotes for FSC Forest Management certification from (other) certification bodies, could you share the fees they proposed?

Please fill in all fields you can answer. I am sure you can easily find the quotes in your emails or files.

Please, exclude the Annual Accreditation Fee (AAF).

	Certification body	Year of the price quote	Total price for 1st year of (re)certification	Total price for 2nd year after certification	Currency
Price quote 1	Please choose...	Please choose...			Please choose...
Price quote 2	Please choose...	Please choose...			Please choose...
Price quote 3	Please choose...	Please choose...			Please choose...

Question help

Any comment?

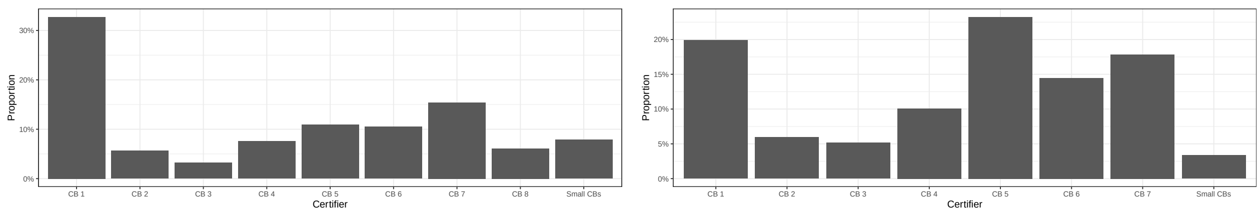
***Would you allow me to share your responses with other scientists who adhere to the data policy of this survey?**

Yes | No

Second, respondents tend to have been certified for a slightly longer time than the whole population as Figures A17 show. This might be related to earlier certified FMUs being more intrinsically motivated for sustainability and more motivated to contribute to corresponding research and a survey. This is a possibility that will need to be considered in the interpretation of the results of my study later on.

Third, there are relatively more respondents from Western Europe and relatively less from Eastern Europe. Region dummies will help to control for this in the price estimation in the main project.

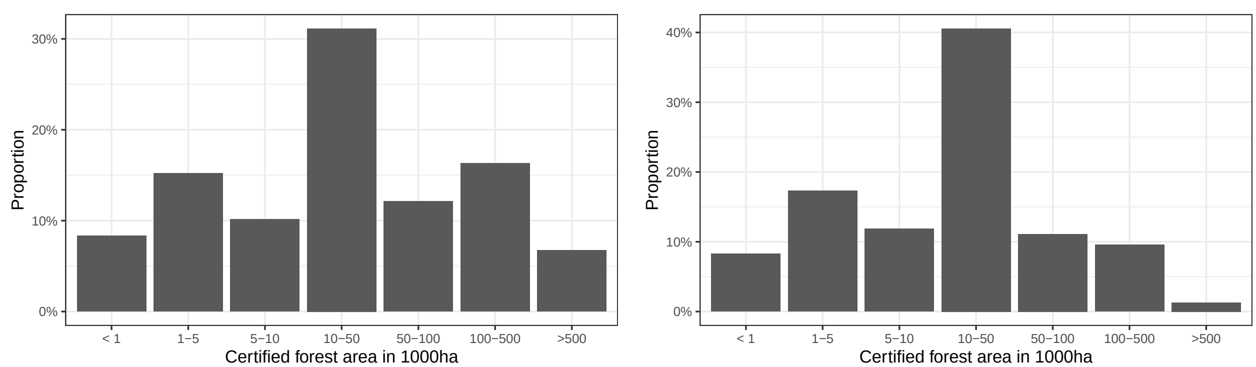
Figure A15: Distribution across certifiers



(a) Target population: FSC FM certificate holders, valid in June 2020

(b) Survey respondents

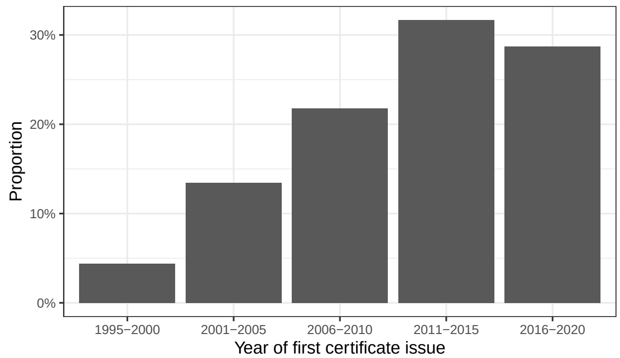
Figure A16: Distribution across forest size categories



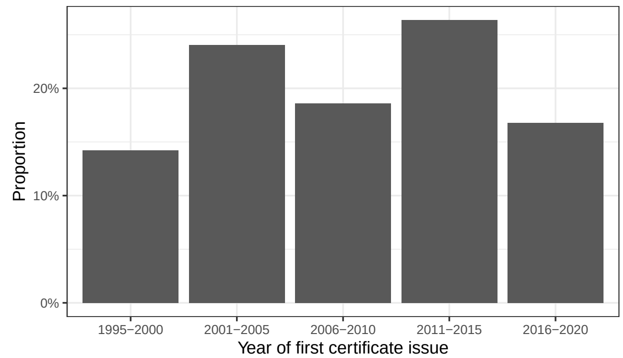
(a) Target population: FSC FM certificate holders, valid in June 2020

(b) Survey respondents

Figure A17: Distribution across initial certification year categories

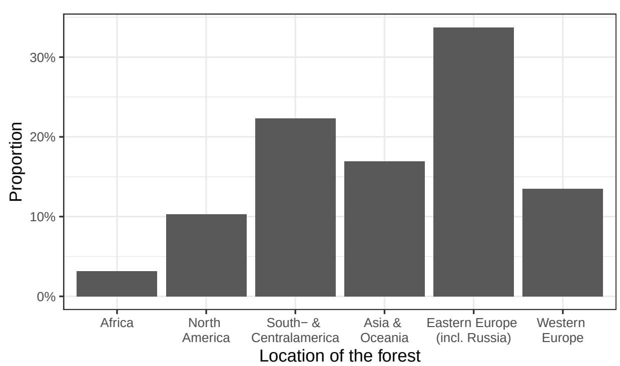


(a) Target population: FSC FM certificate holders, valid in June 2020

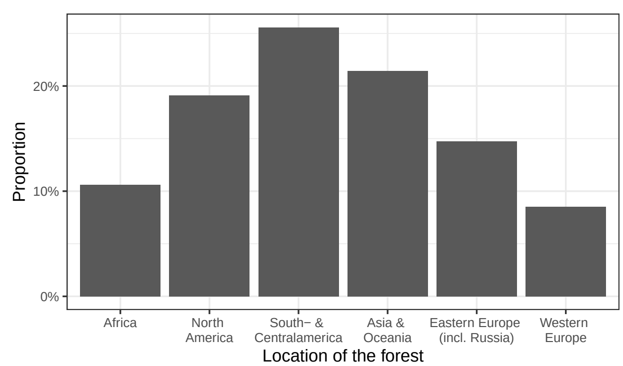


(b) Survey respondents

Figure A18: Distribution across regions



(a) Target population: FSC FM certificate holders, valid in June 2020



(b) Survey respondents

	Forest location	Respondents
1	Africa: - Eastern Africa	3
2	Africa: - Middle Africa	3
3	Africa: - Southern Africa	9
4	Africa: - Western Africa	1
5	Americas: - Caribbean	1
6	Americas: - Central America	17
7	Americas: - Northern America	31
8	Americas: - South America	49
9	Asia: - Central Asia	2
10	Asia: - Eastern Asia	18
11	Asia: - South-eastern Asia	23
12	Asia: - Southern Asia	3
13	Europe: - Eastern Europe (including Northern Asia)	53
14	Europe: - Northern Europe	24
15	Europe: - Southern Europe	23
16	Europe: - Western Europe	45
17	Oceania: - Australia and New Zealand	9
18	Oceania: - Melanesia, Micronesia & Polynesia	2
19	No answer	11

B.3.3 Prediction of unobserved prices

B.3.3.1 Accounting for selection into the survey

1. Following Heckman (1979), I model survey participation as the outcome of a Probit model, where the utility of survey participation is a function of predictors that are selected using repeated cross-validation with Lasso regularization:

$$u\_SurveyParticipation_{ijt} = \underbrace{\boldsymbol{f}'_{it}\tilde{\boldsymbol{\rho}}^{\boldsymbol{fs}} + \boldsymbol{x}'_{jt}\tilde{\boldsymbol{\rho}}^{\boldsymbol{xs}}}_{E[u\_SurveyParticipation_{ijt}|boldsymbol f_{it},\boldsymbol{x_{jt}}]} + v_{ijt}$$
$$SurveyParticipation_{ijt} = 1\{u\_SurveyParticipation_{ijt} > 0\} \tag{24}$$

2. I assume that the unobserved shocks to survey participation and prices, v_{ijt} and ϵ^p_{ijt} are jointly normal, with covariance σ_{21}. This implies that conditional mean of the price in the selected sample can then be written as:

$$\begin{aligned} &E\left[\log\left(\frac{p_{ijt}}{\log(\overline{area}_t)}\right)\Big|\boldsymbol{f_{it}}, \boldsymbol{x_{jt}}, SurveyPart_{ijt} = 1\right] \\ &= E\left[\log\left(\frac{p_{ijt}}{\log(\overline{area}_t)}\right)\Big|\boldsymbol{f_{it}}, \boldsymbol{x_{jt}}\right] + E\left[\epsilon^p_{ijt}|v_{ijt} > -\mathbf{f}'_{ijt}\tilde{\rho}\right] \\ &= E\left[\log\left(\frac{p_{ijt}}{\log(\overline{area}_t)}\right)\Big|\boldsymbol{f_{it}}, \boldsymbol{x_{jt}}\right] + \sigma_{21} InverseMillsRatio(\boldsymbol{f}'_{it}\tilde{\boldsymbol{\rho}}^{\boldsymbol{fs}} + \boldsymbol{x}'_{jt}\tilde{\boldsymbol{\rho}}^{\boldsymbol{xs}}) \end{aligned} \tag{25}$$

   Not accounting for the second term would lead to omitted variable bias.

3. I construct the binary variable $SurveyParticipation_{ijt}$ for survey participation in the audit data. Given that the survey data is anonymized, the price quotes cannot be matched one by one to FMUs in the audit data. I therefore construct types of FMUs based on the characteristics collected in the survey and available in the FMU panel. These are the certifier, the year, the year of initial certification, the forest size, whether it has a group rather than an individual certificate, the region (defined as in the market) and whether the year requires a recertification rather than an annual surveillance audit.[73]
4. I estimate $\tilde{\rho}$ by probit regression of $SurveyParticipation_{ijt}$ and construct the Inverse Mills Ratio $InverseMillsRatio(\boldsymbol{f}'_{it}\tilde{\boldsymbol{\rho}}^{\boldsymbol{fs}} + \boldsymbol{x}'_{jt}\tilde{\boldsymbol{\rho}}^{\boldsymbol{xs}})$.
5. I add the Inverse Mills Ratio as an additional regressor to the pricing model, but do

[73] The year of initial certification and the forest size are given in categories in the survey data, as described in the data section of this paper. In the models of survey participation and pricing, I use the average by category and market in the FMU panel as a numerical value.

Table A4: Log of annual certification fee in 1000 USD per log area in 1000 ha, PPI adj.

	(1)	(2)	(3)
Certified area in 1000ha (lower bound by cat.) x log_country is landlocked_std	-0.061^{**} (0.028)	-0.059^{**} (0.029)	-0.075^{***} (0.025)
Internal dist. of country x log_country is landlocked_std	0.105^{***} (0.027)	0.105^{***} (0.027)	0.098^{***} (0.024)
Internal dist. of country x (Re-)Certification audit_std	0.073 (0.075)	0.075 (0.076)	0.077 (0.093)
No. of certified downstream firms in country in 2023 x (Re-)Certification audit_std	0.073 (0.044)	0.072 (0.045)	0.078^{*} (0.046)
% of privately owned forest in country in 2016 x nat. avg. Longitude_std	0.229^{***} (0.062)	0.224^{***} (0.063)	0.241^{***} (0.062)
% of privately owned forest in country in 2016 x log_% of publicly owned forest in country in 2015_std	0.032 (0.066)	0.032 (0.065)	0.036 (0.067)
Forest nat. total profit tax (in %)_X___Forest nat. total profit tax (in %)sq_std	0.090 (0.079)	0.098 (0.079)	0.055 (0.085)
Forest nat. total profit tax (in %)_X___log_Yrs. with FSC cert._std	-0.111^{*} (0.065)	-0.122^{*} (0.070)	−0.106 (0.073)
Exported share of nat. production of sawnwood x nat. export from wood chips etc. in m3_std	−0.018 (0.071)	−0.021 (0.073)	−0.015 (0.073)
Exported share of nat. production of sawnwood x No. of promotional license holders in country by year_std	0.008 (0.040)	0.007 (0.040)	0.007 (0.041)
Exported share of nat. production of sawnwood x log_nat. import from fibre furnish in t_std	0.073 (0.055)	0.079 (0.058)	0.077 (0.059)
Exported share of nat. production of woodbased panels x log_Certified area in 1000ha (lower bound by cat.)_std	-0.216^{***} (0.063)	-0.222^{***} (0.064)	-0.211^{***} (0.065)
Exported share of nat. production of woodbased panels x (Re-)Certification audit_std	0.071^{**} (0.028)	0.074^{**} (0.030)	0.081 (0.073)
nat. export from fibreboard in m3 x Group certificate (vs. individual)_std	−0.021 (0.048)	−0.020 (0.048)	−0.023 (0.050)
nat. export from woodbased panels in m3 x No. of certified downstream firms in country in 2023 sq._std	0.042 (0.062)	0.037 (0.063)	0.020 (0.065)
nat. avg. latitude x log_Certified area in 1000ha (lower bound by cat.)_std	-0.199^{**} (0.083)	-0.196^{**} (0.084)	-0.197^{**} (0.086)
nat. avg. latitude x log_country has an FSC national standard, by year_std_std	−0.012 (0.064)	−0.012 (0.064)	−0.004 (0.068)
nat. avg. Longitude x nat. net export from woodbased panels in 1000 USD_std	−0.041 (0.063)	−0.051 (0.065)	−0.031 (0.066)
nat. avg. Longitude x log_Certified area in 1000ha (lower bound by cat.)_std	-0.337^{***} (0.097)	-0.337^{***} (0.097)	-0.352^{***} (0.101)
nat. avg. Longitude x log_country is landlocked_std	0.040 (0.032)	0.042 (0.032)	0.051^{*} (0.029)
nat. avg. Longitude x Classified as plantation_std	0.143^{***} (0.047)	0.140^{***} (0.046)	0.139^{***} (0.047)
Yrs. with FSC cert. sq. x Exported share of nat. production of sawnwood sq._std	-0.057^{***} (0.020)	-0.062^{***} (0.023)	-0.060^{***} (0.023)
Yrs. with FSC cert. sq. x nat. avg. latitude sq._std	−0.020 (0.056)	−0.020 (0.056)	−0.021 (0.057)
Yrs. with FSC cert. sq. x log_country has an FSC national standard, by year_std_std	−0.023 (0.047)	−0.026 (0.049)	−0.026 (0.049)
% of publicly owned forest in country in 2015 sq. x log_country has an FSC national standard, by year_std_std	-0.112^{***} (0.040)	-0.107^{**} (0.042)	-0.108^{**} (0.044)
Forest nat. total profit tax (in %)sq x (Re-)Certification audit_std	−0.011 (0.042)	−0.014 (0.043)	−0.004 (0.045)
Forest nat. total profit tax (in %)sq x Classified as plantation_std	0.034 (0.068)	0.041 (0.072)	0.068 (0.082)
Exported share of nat. production of woodbased panels sq. x nat. import from wood pulp in t sq._std	0.083 (0.065)	0.083 (0.066)	0.082 (0.068)
Exported share of nat. production of woodbased panels sq. x No. of promotional license holders in country by year sq._std	0.019 (0.045)	0.021 (0.046)	0.013 (0.047)
nat. export from fibreboard in m3 sq. x log_Exported share of nat. production of industrial roundwood_std	0.125^{***} (0.037)	0.129^{***} (0.039)	0.140^{***} (0.042)
Years since first FSC national standard_std sq. x nat. net export from wood chips etc. in 1000 USD sq._std	0.104^{**} (0.053)	0.103^{*} (0.053)	0.101^{*} (0.054)
nat. net export from woodbased panels in 1000 USD sq. x (Re-)Certification audit_std	0.009 (0.030)	0.010 (0.030)	0.007 (0.036)
nat. net export from wood chips etc. in 1000 USD sq. x log_Certified area in 1000ha (lower bound by cat.)_std	-0.100^{**} (0.050)	-0.101^{**} (0.049)	-0.094^{*} (0.051)
log_Certified area in 1000ha (lower bound by cat.) x log_% of privately owned forest in country in 2016_std	0.013 (0.074)	0.021 (0.074)	0.015 (0.073)
log_Certified area in 1000ha (lower bound by cat.) x log_Forest nat. total profit tax (in %)_std	−0.020 (0.066)	−0.028 (0.067)	−0.024 (0.067)
log_nat. export from wood chips etc. in m3 x (Re-)Certification audit_std	−0.001 (0.051)	−0.004 (0.052)	−0.012 (0.087)
log_nat. export from wood chips etc. in m3 x Group certificate (vs. individual)_std	0.050 (0.053)	0.049 (0.053)	0.061 (0.056)
log_country has an FSC national standard, by year_std x (Re-)Certification audit_std	−0.013 (0.050)	−0.013 (0.051)	−0.020 (0.051)
Group certificate (vs. individual) x Classified as plantation_std	0.063 (0.049)	0.061 (0.049)	0.058 (0.051)
Expected relative rigor x log_nat. export from fibreboard in 1000 USD per m3 (t-avg.)_std	0.025 (0.036)	0.026 (0.036)	0.043 (0.036)
Certifier market_group_no_sales_inclNone x nat. avg. latitude (t-avg.)_std	−0.009 (0.053)	−0.007 (0.054)	0.004 (0.055)
nat. export of wood chips in 1000 USD per cubic metre (t-avg.) x nat. export from wood chips etc. in m3 (t-avg.) sq._std	0.025 (0.029)	0.024 (0.029)	0.028 (0.029)
nat. export from paper products in 1000 USD per t (t-avg.) x log_ch_country_yr_Export_t_Wood_pulp (t-avg.)_std	0.054^{*} (0.032)	0.056^{*} (0.033)	0.055^{*} (0.033)
Local currency unit per USD (t-avg.) x Country with certifier's closest office and country share the language (t-avg.) sq._std	0.054^{*} (0.029)	0.054^{*} (0.029)	0.059^{*} (0.032)
nat. avg. latitude (t-avg.) x log_country is landlocked (t-avg.)_std	0.078 (0.063)	0.071 (0.065)	0.064 (0.067)
nat. avg. Longitude (t-avg.) x Certifier's market share among FSC certifiers sq._std	0.106^{*} (0.061)	0.102^{*} (0.061)	0.111^{*} (0.062)
Certifier's years with accreditation sq. x country is landlocked (t-avg.) sq.	−0.004 (0.008)	−0.006 (0.009)	−0.004 (0.009)
avg. dist. to certifier's headquarter by market in 1000 km sq. x country is landlocked (t-avg.) sq._std	-0.160^{***} (0.062)	-0.147^{**} (0.068)	-0.143^{**} (0.071)
avg. dist. to certifier's headquarter by market in 1000 km sq. x nat. import from fibre furnish in t (t-avg.) sq._std	0.031 (0.044)	0.033 (0.045)	0.040 (0.047)
Expected relative rigor sq. x Exported share of nat. production of woodbased panels (t-avg.) sq._std	−0.035 (0.042)	−0.036 (0.042)	−0.025 (0.046)
Certifier market_group_no_sales_inclNone sq. x log_nat. export of wood chips in 1000 USD per cubic metre (t-avg.)_std	−0.022 (0.033)	−0.024 (0.033)	−0.024 (0.033)
log_Log of nat. score in Corruption Perceptions Index_fsc (t-avg.) x log_dist. to the certifier's closest office (t-avg.)_std	0.047 (0.054)	0.041 (0.055)	0.035 (0.056)
log_nat. export from woodbased panels in 1000 USD per m3 (t-avg.) x log_ch_country_yr_Export_t_Wood_pulp (t-avg.)_std	−0.005 (0.039)	−0.006 (0.039)	−0.002 (0.040)
Certified area in 1000ha (lower bound by cat.) (t-avg.) x _year2018_std	−0.041 (0.042)	−0.044 (0.042)	−0.045 (0.044)
Certified area in 1000ha (lower bound by cat.) (t-avg.) x cb_Certifier 5_std	0.008 (0.054)	0.010 (0.053)	0.012 (0.055)
nat. import from fibre furnish in t (t-avg.) x _year2019_std	0.003 (0.042)	−0.002 (0.043)	−0.004 (0.044)
Country with certifier's closest office and country share the language (t-avg.) x cb_Certifier 5_std	-0.113^{**} (0.057)	-0.116^{**} (0.057)	-0.112^{*} (0.058)
nat. avg. latitude (t-avg.) x cb_Certifier 6_std	-0.156^{***} (0.044)	-0.160^{***} (0.044)	-0.172^{***} (0.046)
nat. avg. latitude (t-avg.) x cb_Certifier 7_std	0.067 (0.067)	0.065 (0.068)	0.065 (0.069)
nat. avg. Longitude (t-avg.) x cb_Certifier 5_std	0.040 (0.048)	0.039 (0.048)	0.040 (0.048)
avg. dist. to certifier's headquarter by market in 1000 km sq. x ___groupEastern_and_Southern_Africa___all_same_std	−0.014 (0.055)	−0.022 (0.058)	−0.024 (0.058)
country is landlocked (t-avg.) sq. x cb_Certifier 5_std	−0.042 (0.046)	−0.033 (0.050)	−0.035 (0.052)
log_Relative rigor (certifier constant) x cb_Certifier 3_std	-0.057^{**} (0.024)	-0.056^{**} (0.024)	-0.084^{***} (0.029)
log_Relative rigor (certifier constant) x cb_Certifier 4_std	0.040 (0.037)	0.043 (0.037)	0.040 (0.036)
log_Log of nat. score in Corruption Perceptions Index_fsc (t-avg.) x _year2019_std	0.016 (0.033)	0.013 (0.034)	0.026 (0.034)
log_nat. export from roundwood in 1000 USD per m3 (t-avg.) x _year2019_std	0.049 (0.033)	0.052 (0.034)	0.048 (0.034)
log_nat. import from paper products in t (t-avg.) x _year2018_std	−0.016 (0.032)	−0.016 (0.032)	−0.018 (0.033)
Control functions:			
Small certifiers			0.055 (0.102)
Certifier 1			−0.081 (0.118)
Certifier 2			0.159 (0.151)
Certifier 3			0.154 (0.136)
Certifier 4			−0.020 (0.157)
Certifier 5			−0.046 (0.127)
Certifier 6			0.066 (0.144)
Certifier 7			−0.122 (0.399)
Small certifiers			
Inverse Mills Ratio from probit of survey part.		−0.058 (0.121)	−0.030 (0.126)
Constant	1.104^{***} (0.037)	1.196^{***} (0.189)	1.130^{***} (0.200)
Observations (choice situations)	379	379	379
R^2	0.772	0.772	0.776
Adjusted R^2	0.722	0.722	0.719
Residual Std. Error	0.410 (df = 311)	0.410 (df = 310)	0.412 (df = 302)
F Statistic	15.679^{***} (df = 67; 311)	15.409^{***} (df = 68; 310)	13.728^{***} (df = 76; 302)

Notes: Standard errors are robust to heteroskedasticity and serial correlation. They do not account for the variance of the generated regressors used as the control function, yet.
Abbreviations: std. (standardized), sq. (squared)., nat. (national), avg. (average), t-avg. (avg. by market), dist. (distance), cat. (category), fun. (function)$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

not include its predictions in the predictions of prices for the audit data.

A positive estimate of σ_{21} would suggest that a positive price shock is associated with a higher shock to the utility of survey participation, and vice versa. In other words, FMUs that pay higher prices are more likely to participate in the survey. This is what I expect since FMUs that are able to pay higher prices than other firms, are likely to have a less restrictive budget constraint for certification, which is likely to be correlated with having time resources to participate in a survey. Another potential driver of a positive association between prices and survey participation is that FMUs' with higher intrinsic motivation for certification are more likely to be willing to pay higher prices for higher quality and are at the same time also more likely to participate in the survey. Such a positive estimate σ_{21} of would suggest that without controlling for selection into the survey, I would overestimate prices, and vice versa. In Table A4, I do find a positive, but insignificant estimate of σ_{21}.

C Appendix: Estimation and results

C.1 Stage 3: Violation reporting

C.1.1 Selection of baseline violation predictors

To select relevant predictors $\boldsymbol{f}_{it}^{v}$ in the model of violation reporting, I take the following steps.

1. I start by including potentially relevant certifier-invariant audit and FMU characteristics, for example certificate types, categories of forest types and tree species, country-level variables on the trade of various wood products, corruption levels, national FSC standards, number of FSC-certified retailers by country, etc.
2. I remove variables with less than 10% correlation with violation reports or a correlation with a p-value of more than 0.5.
3. I include squares and logs of the remaining variables (if numeric) and include all possible interactions (leaving around 32,000 predictors)
4. I remove near-constant and collinear predictors (leaving around 10,800 predictors).
5. I remove predictors with less than 10% correlation with violation reports or a correlation of less tgab 50% with a p-value of more than 0.1 (leaving around 2,300 predictors).
6. I remove predictors that are more than 90% correlated with other predictors (leaving around 300 predictors).

7. I select relevant predictors by estimating model (12) with PPML using LASSO regularization, excluding the certifier variant variables $\boldsymbol{x^v_{ijt}}$ and certifier fixed effects from the penalization. I use a slightly higher penalty factor than suggested by cross-validation since I do not predict much out-of-sample in certifier-invariant characteristics. The in-sample R-Squared, adapted for the count data model following (Cameron and Windmeijer, 1996), is 0.28.

C.1.2 Bias correction for count data models with multinomial selection

This section provides further detail on the derivation of a control function for sample selection in the Stage 3 model.

v_{ijt} are only observed from the certifier j chosen by i in t, i.e. conditional on the choices $y_{ikt} \forall k \in J_t$ and their determinants $\boldsymbol{V_{it}}$:

$$\mathbb{E}[v_{ijt}|\boldsymbol{y_{it}}, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}] = \exp(\mu^v_{ijt}) \sum_{k \in \{1,\ldots,J\}} y_{ikt} \mathbb{E}[\exp(\eta_i)|y_{ikt} = 1, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}] \tag{26}$$

This introduces sample selection bias since $\mathbb{E}[\exp(\eta_i)|\boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, y_{ikt} = 1, \boldsymbol{V_{it}}]$ will not necessarily be constant across the chosen alternatives k. This is the case if η_i is correlated with the differences between the taste shocks $\epsilon^u_{i1t} - \epsilon^u_{ijt}, \ldots; \epsilon^u_{iJt} - \epsilon^u_{ijt}$ from the demand model. A particularly relevant case of that is if firms that are less compliant due to factors not included in $\boldsymbol{f^v_{it}}$ select into less rigorous certifiers.

To account for the sample selection, I derive and include a control function. Following Heckman (1979), the idea is to control for the mean of the unobservable $\exp(\eta_i)$ within the selected sample by specifying the joint distribution of η_i and the unobservable terms of the demand model. The problem differs from the standard control function approach in two ways. First, the choice is multinomial. Second, the model is a count data model so that the conditional mean of $\exp(\eta_i)$ is not additive. To deal with that, I combine the approaches by Lee (1983) for multinomial selection problems and by Terza (1998) for count data models with binary selection. To the best of my knowledge, I am the first to do so.

Following Lee (1983), I rewrite the multinomial selection in terms of maximum order statistics:

$$y_{ijt} = 1 \text{ iff } V_{ijt} \geq e_{ijt} \text{ where } e_{ijt} \equiv max_{k \neq j}(V_{ikt} + \epsilon^u_{ikt} - \epsilon^u_{ijt}) \tag{27}$$

The marginal distribution of e_{ijt} is such that $F_j(V_{ijt}) = s_{ijt}$, the conditional choice probability defined in the paper. e_{ijt} is transformed into a standard normal random variable by defining

$$e^*_{ijt} \equiv G_j(e_{ijt}) \equiv \Phi^{-1}(F_j(e_{ijt})) \tag{28}$$

e^*_{ijt} and η_i are jointly normally distributed with zero means, variances 1 and σ_η and correlation coefficient ρ_j. Since $G_j()$ is a strictly increasing function, (27) and (28) translate into $y_{ijt} = 1$ iff $e^*_{ijt} \leq G_j(V_{ijt})$.

By the Law of Iterated Expectation,

$$\begin{aligned}&\mathbb{E}[\exp(\eta_i)|y_{ikt} = 1, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}]\\ &= \mathbb{E}\Big[\mathbb{E}[\exp(\eta_i)|e^*_{ijt}, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}]|e^*_{ijt} \leq G_j(V_{ijt}), \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}\Big]\end{aligned} \tag{29}$$

Terza (1998) (Appendix A) shows the following for two random variables η_i and e^*_{ijt} that satisfy the same assumptions as in my case (ϵ and v in his paper), in particular a joint normal distribution with zero means and covariance matrix $\Sigma = \begin{bmatrix} \sigma_\eta & \sigma_\eta\rho_j \\ \sigma_\eta\rho_j & 1 \end{bmatrix}$:

$$\mathbb{E}[\exp(\eta_i)|e^*_{ijt}, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}] = \exp(\rho_j\sigma_\eta e^*_{ijt} + \sigma^2_\eta/2(1-\rho^2_j)) \tag{30}$$

Consequently

$$\begin{aligned}&\mathbb{E}\Big[\mathbb{E}[\exp(\eta_i)|e^*_{ijt}, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}]|e^*_{ijt} \leq G_j(V_{ijt}), \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}\Big]\\ &= \exp(\sigma^2_\eta/2(1-\rho^2_j))\mathbb{E}\Big[\exp(\theta_j e^*_{ijt})|e^*_{ijt} \leq G_j(V_{ijt}), \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}\Big]\end{aligned} \tag{31}$$

where $\theta_j \equiv \rho_j\sigma_\eta$

The next steps differ follow equation (9) in Terza (1998), with the difference that my condition is $e^*_{ijt} \leq G_j(V_{ijt})$, exploiting that $\Phi(-x) = 1 - \Phi(x)$:

$$\begin{aligned}
&\mathbb{E}\Big[\exp(\theta_j e^*_{ijt})|e^*_{ijt} \leq G_j(V_{ijt}), \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}\Big] \\
&= \frac{\mathbb{E}\Big[\exp(\theta_j e^*_{ijt}) \wedge e^*_{ijt} \leq G_j(V_{ijt})|\boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}\Big]}{\Phi(G_j(V_{ijt}))} \\
&= \frac{\int_{-\infty}^{G_j(V_{ijt})} \exp(\theta_j e^*_{ijt})\phi(e^*_{ijt})de^*_{ijt}}{\Phi(G_j(V_{ijt}))} \\
&= \frac{\int_{-\infty}^{G_j(V_{ijt})} \exp(\theta_j e^*_{ijt})\exp(-\frac{(e^*_{ijt})^2}{2})de^*_{ijt}}{\sqrt{2\pi}\Phi(G_j(V_{ijt}))} \\
&= \frac{\int_{-\infty}^{G_j(V_{ijt})} \exp(-\frac{(e^*_{ijt}-\theta_j)^2}{2} + \frac{\theta_j^2}{2})de^*_{ijt}}{\sqrt{2\pi}\Phi(G_j(V_{ijt}))} \\
&= \frac{\exp(\frac{\theta_j^2}{2})\int_{-\infty}^{G_j(V_{ijt})} \exp(-\frac{(e^*_{ijt}-\theta_j)^2}{2})de^*_{ijt}}{\sqrt{2\pi}\Phi(G_j(V_{ijt}))} \\
&= \frac{\exp(\frac{\theta_j^2}{2})\sqrt{2\pi}\Phi(G_j(V_{ijt}) - \theta_j)}{\sqrt{2\pi}\Phi(G_j(V_{ijt}))}
\end{aligned} \tag{32}$$

Plugging this into (30), one obtains

$$\mathbb{E}[\exp(\eta_i)|y_{ikt} = 1, \boldsymbol{x^v_{ijt}}, j, \boldsymbol{f^v_{it}}, \boldsymbol{V_{it}}] = \exp(\sigma^2_\eta/2)\frac{\Phi(G_j(V_{ijt}) - \theta_j)}{\Phi(G_j(V_{ijt}))} \tag{33}$$

If $\theta_j = 0$, η_i and e_{ijt} are independent, then the control function $\frac{\Phi(G_j(V_{ijt})-\theta_j)}{\Phi(G_j(V_{ijt}))}$ equals 1 and is thus irrelevant. Otherwise the estimates of $\boldsymbol{\gamma}$ and $\boldsymbol{r}$ would be inconsistent without inclusion of the control function. If θ_j is positive, forestries choosing j have a higher η_i, i.e. more true violations, than the average forestry with the same characteristics, and vice versa. A negative correlation of θ_j with r_j would thus suggest selection on unobservables of less compliant firms tend to select into less rigorous certifiers.

The model with the control function is estimated via Poisson Pseudo Maximum Likelihood as in Egger et al. (2011). Estimates of $G_j(V_{ijt})$ are obtained from the standard normal quantiles of the conditional choice probabilities estimated in a first stage version of the demand model, that replaces violation reports with its predictors.[74]

[74] I estimate a simplified version of the demand model (5), but without including expected relative rigor, baseline violations, and prices explicitly. Certifier-market constants capture variation in expected relative rigor across regions, years, and certifiers. Interactions of certifier dummy variables with observed FMU characteristics capture variation in baseline violations.

C.1.3 Conditional exogeneity of the accreditation body's inspection assignments

FSC's accreditation body assigns audits without rigid rules but taking several factors into account:[75]

1. Their main focus is to inspect where they suspect leniency. The model controls for various certifier-invariant and certifier-variant determinants of leniency. ω_2 identifies the (expected) effect of inspections on violation reporting if the accreditation body does not assign inspections based on additional information about leniency (which FMUs observe as well).
2. They tend to inspect more in regions in which they expect lower compliance by FMUs. My rich set of controls should capture this variation.
3. They inspect natural forests more often than plantations since they expect greater welfare effects of improved violation reporting in those forests. The plantation dummy in the model controls for that.
4. They tend to inspect larger certifiers more often, but also new entrants and certifiers in markets with new entrants. That is, accreditation inspections may vary with certifiers' market shares. That variation with market shares does not cause endogeneity of accreditation inspections, unless it correlates with determinants of (expected) violation reports. Such correlation is precisely what the observable controls and the control function take care of. As noted above, the control function is particularly reliable for across-market variation, including variation in market shares. I am, thus, not concerned about endogeneity in that dimension.
5. Conditional on the mentioned factors, the accreditation body may assign inspections randomly or based on practical considerations that are unlikely to correlate with violation reports, such as the location of their assessors.

The fourth and fifth factor generate exogenous variation in inspection assignments that identifies ω_2.

[75] This information is based on ASI (2021) and an interview with the accreditation body's staff.

C.1.4 Results

Figure A19 presents the main results from the estimation of the model of violation reporting: the implied ranking of certifiers in terms of rigor type $\hat{r}_j$. The figure plots the estimates as well as their 95% confidence interval. The orange dot and line come from estimation with control function that accounts for potential selection bias, the blue line is from the estimation without that control function. The certifiers are ranked according to the point estimate from the estimation with this control function.

Figure A19: Estimates of rigor types from the Stage 3 model, with and without control function to correct for sample selection bias

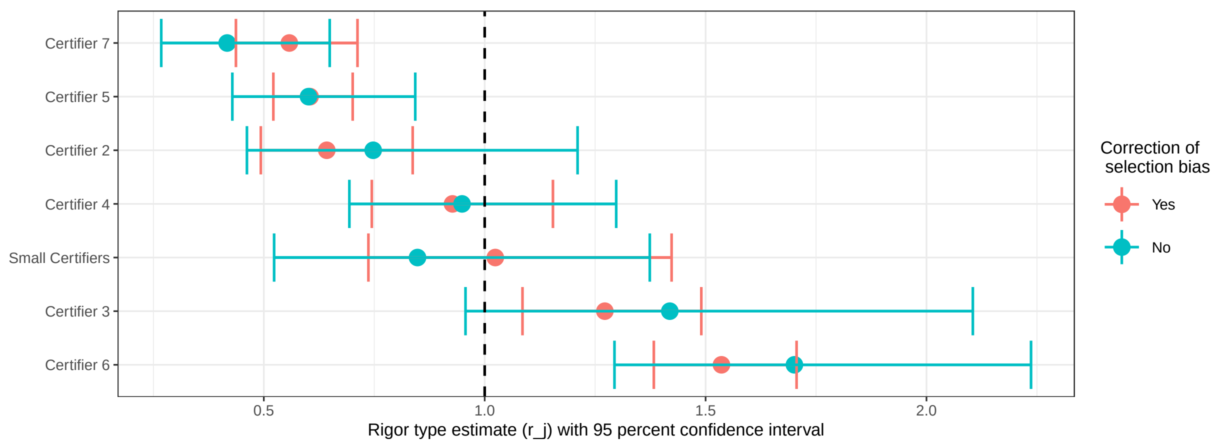


Table A5: No. of violation reports (by year)

	Without control function	*With control function*
	(1)	(2)
Log(r_j), rigor type:		
Certifier 2	−0.291 (0.246)	−0.442*** (0.135)
Certifier 3	0.350* (0.201)	0.241*** (0.081)
Certifier 4	−0.053 (0.160)	−0.076 (0.112)
Certifier 5	−0.509*** (0.172)	−0.503*** (0.075)
Certifier 6	0.531*** (0.140)	0.429*** (0.054)
Certifier 7	−0.875*** (0.226)	−0.584*** (0.125)
Small certifiers	−0.165 (0.246)	0.024 (0.168)
$\boldsymbol{\omega}$:		
Audit inspected by accreditation body	0.776*** (0.156)	0.793*** (0.054)
avg. dist. to certifier's headquarter by market in 1000 km	−0.034** (0.017)	−0.033*** (0.007)
Forest nat. CPI inflation rate_X___nat. net export from roundwood in 1000 USD_std	0.015 (0.093)	0.006 (0.037)
nat. forest area annual net change rate (Percent) x nat. certified forest area (1000 ha)_std	−0.075 (0.132)	−0.080* (0.044)
nat. forest area annual net change rate (Percent) x nat. net export from forest products in 1000 USD sq._std	−0.046 (0.097)	−0.029 (0.026)
nat. forest area annual net change rate (Percent) x log_Certified area in 1000 ha_std	−0.108 (0.081)	−0.111*** (0.031)
nat. avg. Longitude x nat. export from fibre furnish in 1000 USD per t sq._std	0.041 (0.061)	0.048 (0.032)
nat. net export from industrial roundwood in 1000 USD x nat. net export from wood chips etc. in 1000 USD_std	0.214 (0.135)	0.212* (0.119)
nat. net export from paper products in 1000 USD x Forest has gymnosperms (clade of plants)_std	−0.226*** (0.065)	−0.237*** (0.030)
nat. net export from roundwood in 1000 USD x nat. forest area annual net change rate (Percent) sq._std	−0.214** (0.094)	−0.209*** (0.049)
nat. net export from wood chips etc. in 1000 USD x log_Years since first FSC national standard_std_std	−0.082 (0.097)	−0.090** (0.041)
nat. net export from forest products in 1000 USD x nat. export from industrial roundwood in 1000 USD per m3_std	0.034 (0.118)	0.048 (0.038)
Forest nat. total profit tax (in %)_X___nat. avg. latitude sq._std	−0.100 (0.114)	−0.114* (0.058)
Forest nat. total profit tax (in %)_X___log_Yrs. with FSC cert._std	−0.101 (0.090)	−0.091** (0.036)
Above-ground biomass stock in nat. forests (t per ha) x log_Yrs. with FSC cert._std	−0.037 (0.105)	−0.024 (0.037)
No. of certified downstream firms in country in 2023 x log_Yrs. with FSC cert._std	−0.094 (0.092)	−0.092* (0.048)
No. of promotional license holders in country by year x nat. net export from sawnwood in 1000 USD sq._std	−0.108 (0.098)	−0.118** (0.055)
nat. export from woodbased panels in 1000 USD per m3 x Forest has gymnosperms (clade of plants)_std	−0.232*** (0.047)	−0.226*** (0.024)
Exported share of nat. production of industrial roundwood x log_No. of certified public projects in country by year_std	−0.159 (0.156)	−0.156* (0.094)
Forest nat. CPI inflation ratesq x nat. net export from wood chips etc. in 1000 USD sq._std	0.199*** (0.064)	0.205*** (0.052)
nat. avg. latitude sq. x nat. export from sawnwood in 1000 USD per m3 sq._std	−0.055 (0.101)	−0.047 (0.048)
Certified area in 1000 ha sq. x Local currency unit per USD sq._std	−0.028 (0.024)	−0.027*** (0.004)
nat. score in Corruption Perceptions Index sq. x Forest nat. total profit tax (in %)sq_std	0.012 (0.106)	−0.014 (0.052)
nat. GDP PC, PPP (2017 1K USD) sq. x Forest nat. total profit tax (in %)sq_std	−0.024 (0.178)	0.024 (0.093)
nat. GDP PC, PPP (2017 1K USD) sq. x log_Local currency unit per USD_std	−0.119 (0.088)	−0.114** (0.055)
Fraction of certified forest's products that is rough wood sq. x log_nat. export from sawnwood in 1000 USD per m3_std	−0.131** (0.053)	−0.128*** (0.021)
No. of certified public projects in country by year sq. x Exported share of nat. production of woodbased panels sq._std	−0.071 (0.115)	−0.066* (0.039)
nat. export from roundwood in 1000 USD per m3 sq. x log_No. of certified public projects in country by year_std	−0.052 (0.056)	−0.052 (0.037)
nat. export from sawnwood in 1000 USD per m3 sq. x log_Certified area in 1000 ha_std	0.119 (0.085)	0.117*** (0.027)
log_Certified area in 1000 ha x log_Local currency unit per USD_std	0.024 (0.087)	0.026 (0.032)
log_Yrs. with FSC cert. x log_% of publicly owned forest in country in 2015_std	−0.015 (0.106)	−0.004 (0.040)
log_Yrs. with FSC cert. x Forest has angiosperms (clade of plants)_std	−0.035 (0.074)	−0.022 (0.039)
log_Forest nat. total profit tax (in %) x log_nat. certified forest area (1000 ha)_std	−0.031 (0.113)	−0.020 (0.040)
log_No. of certificate members x log_nat. export from roundwood in 1000 USD per m3_std	0.072 (0.045)	0.076*** (0.019)
Forest has gymnosperms (clade of plants) x Certified forest has a website_std	0.184*** (0.051)	0.179*** (0.027)
nat. forest area annual net change rate (Percent) x _ch_forest_typeNatural_std	−0.104** (0.048)	−0.100*** (0.021)
No of certified forest's products x _cMexico_std	0.019 (0.056)	0.012 (0.018)
No. of promotional license holders in country by year x _cMexico_std	0.078 (0.050)	0.078*** (0.015)
No. of certificate members sq. x _cMexico_std	0.046*** (0.010)	0.047*** (0.003)
% of publicly owned forest in country in 2015 sq. x _ch_forest_typeNatural_std	0.097* (0.053)	0.097*** (0.017)
log_Local currency unit per USD x _ch_forest_typeNatural_std	−0.029 (0.063)	−0.027 (0.027)
log_No. of certificate members x _Southern_and_Southeastern_Asia_std	−0.016 (0.089)	−0.019 (0.014)
log_No of certified forest's products x _Central_America_std	0.040 (0.056)	0.041 (0.030)
(Re-)Certification audit x _ch_forest_typeNatural_std	0.144*** (0.041)	0.118*** (0.013)
θ_j from control functions:		
Certifier 1		−0.061 (0.040)
Certifier 2		−0.256*** (0.091)
Certifier 3		−0.222*** (0.057)
Certifier 4		−0.083 (0.064)
Certifier 5		−0.080 (0.054)
Certifier 6		−0.228*** (0.033)
Certifier 7		0.545*** (0.174)
Small certifiers		0.270 (0.219)
Constant	−0.700*** (0.161)	−0.715*** (0.065)
Observations (choice situations)	3,810	3,810

Notes: Standard errors are robust to heteroskedasticity and serial correlation. They do not account for the variance of the generated regressors used as the control function, yet. Abbreviations: std. (standardized), sq. (squared)., nat. (national), avg. (average), t-avg. (avg. by market), dist. (distance), cat. (category), fun. (function)$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

C.1.5 Discussion: Potential explanations of certifiers' rigor differences

C.1.5.1 Summary Why should certifiers have different rigor types? Different ownership types and intrinsic motivation of managers or staff might explain variation in types. Intrinsically motivated staff might also be willing to work for a lower salary than staff that is not motivated by their activity's cause (Preston, 1989). Certifiers with more intrinsically motivated staff might, thereby, be more cost-efficient for the same level of rigor. In this way, such certifiers could survive despite firms' disutility of higher rigor by offering lower prices.

However, different rigor types can appear and survive even if the rigor type is a strategic long-run choice of for-profit certifiers without intrinsic motivation. This paper's model nests this possibility, even though I do not model it explicitly. The next section discusses this in more detail. Intuitively, a profit-maximizing certifier sets its rigor type by equalizing the marginal cost and the marginal benefits of increasing the rigor type. The cost of increasing rigor can come from increased audit costs and from reduced revenues if FMUs prefer leniency. The size of the reduction in revenues may vary across certifiers due to differentiation in other characteristics, potentially explaining heterogeneity in rigor types. Certifiers' benefits from increasing rigor arise from a reduced likelihood of accreditation suspension, which increases the expected future profits. This fact can rationalize the heterogeneity in rigor types due to heterogeneity in profits. Classic reputation theory suggests that larger certifiers with higher profits have reason to be more concerned about losing market access and may, thus, be more rigorous than smaller certifiers (Klein and Leffler, 1981). However, such a pattern is not found here. Certifiers 2 and 7 are very large certifiers both in FSC and other markets. They are the only two publicly traded FSC certifiers. The overall correlation of certification companies' total revenue in 2020 with the relative rigor estimates is negative and statistically significant. Given the low number of certifiers, this finding cannot be interpreted as a pattern in the opposite direction, but, rather, as motivation for further research. An effect of rigor on the likelihood of accreditation suspensions can still rationalize the heterogeneity of rigor if certifiers differ in the degrees to which they discount future profits.

C.1.5.2 Rigor as a strategic choice This paper does not consider rigor a strategic variable that certifiers can choose freely in each market. In particular, the model treats the effect of accreditation inspections and the loss of quality control with further distance from the headquarters as deterministic. It fixes other differences in rigor between certifiers to a ratio r_j. Nevertheless, the model allows some discussion of certifiers' potential strategic choice of r_j.

In each market, the profit the certifier expects is $\mathbb{E}[\pi_{jt}|c_{jt}, \bar{c}_t, s_{jt}, j \in J_t]$, defined in model 8, if they are accredited. A marginal decrease in r_j would raise that conditionally expected profit in each market through an effect on revenues and an effect on the marginal cost of certification:

$$\frac{\partial \mathbb{E}[\pi_{jt}|c_{jt}, \bar{c}_t, s_{jt}, j \in J_t]}{\partial r_j} = N_t \bar{c}_t \left[\underbrace{\left(\frac{\partial k_{jt}}{\partial r_j} - 1 \right) c_{jt} s_{jt} + (k_{jt} - 1)\, c_{jt} \frac{\partial s_{jt}}{\partial r_j}}_{\text{Effect on revenues}} + \underbrace{(k_{jt} - 1) \frac{\partial c_{jt}}{\partial r_j} s_{jt}}_{\text{Effect on cost}} \right] \tag{34}$$

where $s_{jt} = s_{jt}(\boldsymbol{r})$ and $k_{jt} = k_{jt}\Big(s_{jt}(\boldsymbol{r}), s_{jt|FSC}(\boldsymbol{r})\Big)$. I expect reduced rigor to decrease the marginal cost of certification but do not quantify that effect due to a lack of identifying variation. But the demand model estimates allow quantifying the main marginal effect on revenues. Table A6 shows the increase in conditionally expected profits due to a decrease in the rigor type r_j by one percentage point. It shows that the willingness to pay for leniency creates strong incentives for certifiers to lower their rigor.

Table A6: Certifiers' marginal revenue benefits from leniency

Statistic	Median	Mean	Min	Max
Total	432,641.80	747,577.80	118,426.50	1,793,817.00
By market	82.81	148.11	4.09	791.72

Notes: Predicted reduction in conditional expected profits due to demand effects of an increase of the rigor type by 1 p.p. in 1000 USD

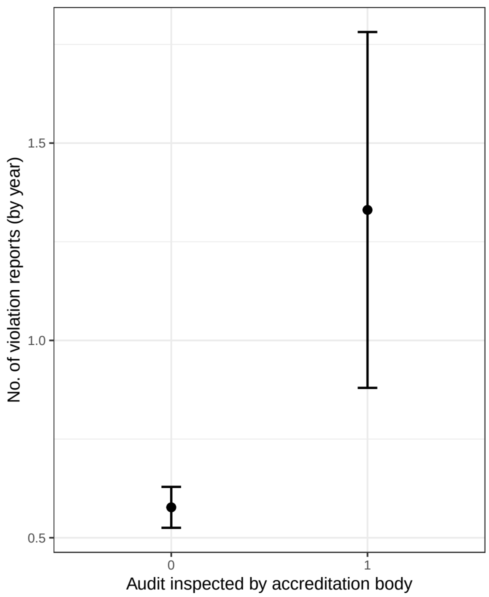
If certifiers choose their rigor types strategically, the cost of decreasing the rigor type below the levels identified as optimal must compensate for the benefits. First, an increased likelihood of accreditation suspension could cause such a cost of leniency. That alone could rationalize the heterogeneity in rigor across certifiers due to heterogeneity in profits. Classic reputation theory suggests that larger certifiers that profit more have reason to be more concerned about losing market access and may thus be more rigorous (Klein and Leffler, 1981). However, there is no pattern in that sense here. Certifiers 2 and 7 are very large certifiers both in FSC and other markets. They are the only two publicly traded certifiers. The overall correlation of certification companies' total revenue in 2020 with the relative rigor estimates is negative and statistically significant. Given the few certifiers, one cannot interpret this as a pattern in the opposite direction, but it encourages further research. An effect of rigor on the likelihood of accreditation suspensions could still rationalize the heterogeneity in rigor if certifiers differ in the degree to which they discount future profits. A second potential explanation for differences in rigor is that certifier staff that is more intrinsically motivated might also be willing to work for a lower salary than staff that is not motivated by their activity's cause (Preston, 1989). Certifiers with more intrinsically motivated staff might thereby be more cost-efficient for the same level of rigor so that they can survive despite firms' disutility of higher rigor by offering lower prices and reducing risks of losing market access due to low rigor in the future. In line with this hypothesis, the certifiers with lower costs according to the supply estimates also tend to be more rigorous.

C.2 Stage 2: Demand

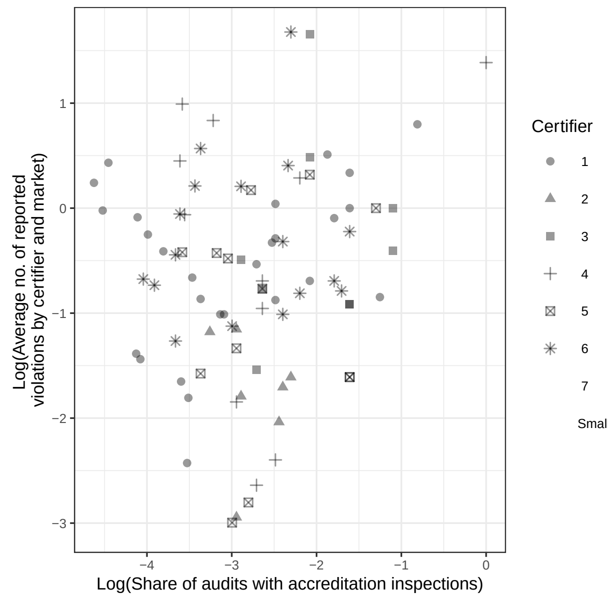
C.2.1 Identifying variation in the likelihood of accreditation inspections

To identify FMUs' willingness to pay for leniency, this paper exploits variation in the share of audits inspected by the accreditation body, by certifier and market, henceforth the *likelihood of accreditation inspections.* Figure A20 (a) shows that the average number of violation reports in audits with accreditation inspections is significantly higher than without. Targeting high-risk markets and FMUs might confound that effect. The empirical model aims to control for that. Figure A20 (b) plots the likelihood of accreditation inspections in logs for positive observations. The plot shows substantial variation across certifiers and markets, which I exploit in the empirical model.

Figure A20: Association of average violation reports with accreditation inspections



(a) Across audits with and without accreditation inspection (Means with 95 percent confidence interval)



(b) Across certifier-market shares of accreditation inspections for non-zero observations

C.2.2 Coefficient tables

Table A7: Preference estimates at the certifier-market-level

	OLS	*2SLS*
	(1)	(2)
$\bar{\alpha}$: Price at certifier-market-level in 1K USD	-0.069^{***} (0.003)	-0.770^{***} (0.178)
$\bar{\beta}^r$: Expected relative rigor	-1.367^{***} (0.046)	-17.366^{***} (1.132)
$\boldsymbol{\beta}^x$:		
First year certifier is available	0.294^{***} (0.019)	1.698^{***} (0.195)
Average distance to certifier's headquarter by market in 1000 km	-0.025^{***} (0.001)	-0.139^{***} (0.016)
$\boldsymbol{\xi}$:		
Certifier 2	-1.274^{***} (0.015)	-3.492^{***} (0.130)
Certifier 3	-0.483^{***} (0.012)	0.709^{***} (0.089)
Certifier 4	-0.818^{***} (0.011)	-0.624^{***} (0.139)
Certifier 5	-0.928^{***} (0.012)	-3.756^{***} (0.210)
Certifier 6	-0.621^{***} (0.013)	3.398^{***} (0.415)
Certifier 7	-1.063^{***} (0.011)	-3.476^{***} (0.147)
Certifier Small Certifiers	-0.326^{***} (0.016)	0.683^{***} (0.151)
FSC x Brazil x 2015	0.101^{***} (0.014)	-0.427^{***} (0.099)
FSC x Brazil x 2016	0.139^{***} (0.016)	-0.425^{***} (0.138)
FSC x Brazil x 2017	-0.502^{***} (0.024)	-1.240^{***} (0.135)
FSC x Brazil x 2018	0.364^{***} (0.018)	-0.150 (0.142)
FSC x Brazil x 2019	0.359^{***} (0.036)	0.530^{***} (0.060)
FSC x Central_America x 2015	-0.115^{***} (0.016)	-0.727^{***} (0.192)
FSC x Central_America x 2016	0.436^{***} (0.020)	-0.162 (0.154)
FSC x Central_America x 2017	-0.055^{**} (0.024)	-0.626^{***} (0.169)
FSC x Central_America x 2018	0.949^{***} (0.015)	0.011 (0.236)
FSC x Central_America x 2019	0.390^{***} (0.023)	-0.167 (0.141)
FSC x Eastern_Asia_excl_China x 2015	-0.176^{***} (0.046)	-0.468^{***} (0.116)
FSC x Eastern_Asia_excl_China x 2016	0.243^{***} (0.025)	-0.010 (0.106)
FSC x Eastern_Asia_excl_China x 2017	1.105^{***} (0.029)	0.826^{***} (0.104)
FSC x Eastern_Asia_excl_China x 2018	0.707^{***} (0.029)	1.858^{***} (0.222)
FSC x Eastern_Asia_excl_China x 2019	-0.914^{***} (0.018)	-0.654^{***} (0.146)
FSC x Eastern_Europe_excl_CIS x 2015	-0.616^{***} (0.012)	0.937^{***} (0.191)
FSC x Eastern_Europe_excl_CIS x 2016	0.224^{***} (0.026)	0.195^{***} (0.051)
FSC x Eastern_Europe_excl_CIS x 2017	0.146^{***} (0.024)	0.296^{***} (0.064)
FSC x Eastern_Europe_excl_CIS x 2018	0.481^{***} (0.020)	0.023 (0.068)
FSC x Eastern_Europe_excl_CIS x 2019	0.223^{***} (0.020)	0.509^{***} (0.079)
FSC x Eastern_and_Southern_Africa x 2015	0.668^{***} (0.029)	-0.613^{***} (0.200)
FSC x Eastern_and_Southern_Africa x 2016	-0.199^{***} (0.034)	-1.035^{***} (0.152)
FSC x Eastern_and_Southern_Africa x 2017	-0.493^{***} (0.025)	-1.303^{***} (0.157)
FSC x Eastern_and_Southern_Africa x 2018	0.720^{***} (0.042)	0.150 (0.188)
FSC x Eastern_and_Southern_Africa x 2019	1.040^{***} (0.040)	0.395^{***} (0.124)
FSC x Northern_America x 2015	-0.184^{***} (0.023)	-0.147^{***} (0.045)
FSC x Northern_America x 2016	-0.778^{***} (0.017)	-0.865^{***} (0.055)
FSC x Northern_America x 2017	-1.032^{***} (0.015)	-1.097^{***} (0.045)
FSC x Northern_America x 2018	-0.654^{***} (0.020)	-0.916^{***} (0.138)
FSC x Northern_America x 2019	-0.970^{***} (0.016)	-0.716^{***} (0.042)
FSC x Oceania x 2015	-0.434^{***} (0.034)	1.115^{***} (0.320)
FSC x Oceania x 2016	0.679^{***} (0.033)	1.022^{***} (0.085)
FSC x Oceania x 2017	-0.476^{***} (0.051)	0.157^{*} (0.082)
FSC x Oceania x 2018	0.637^{***} (0.036)	0.863^{***} (0.082)
FSC x Oceania x 2019	0.323^{***} (0.042)	1.170^{***} (0.192)
FSC x South_America_excl_Brazil x 2015	-0.162^{***} (0.015)	-0.916^{***} (0.175)
FSC x South_America_excl_Brazil x 2016	-0.104^{***} (0.018)	0.129 (0.137)
FSC x South_America_excl_Brazil x 2017	0.095^{***} (0.018)	-0.424^{**} (0.187)
FSC x South_America_excl_Brazil x 2018	0.228^{***} (0.018)	-0.548^{***} (0.196)
FSC x South_America_excl_Brazil x 2019	0.157^{***} (0.019)	-0.240^{**} (0.112)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2016	-0.904^{***} (0.026)	-1.192^{***} (0.037)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2017	-0.493^{***} (0.026)	0.911^{***} (0.140)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2018	-0.019^{*} (0.011)	0.077^{*} (0.043)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2019	-0.112^{***} (0.019)	0.453^{***} (0.095)
FSC x Southern_and_Southeastern_Asia x 2015	0.234^{***} (0.012)	0.564^{***} (0.067)
FSC x Southern_and_Southeastern_Asia x 2016	0.542^{***} (0.017)	0.987^{***} (0.117)
FSC x Southern_and_Southeastern_Asia x 2017	-0.470^{***} (0.016)	-0.466^{***} (0.040)
FSC x Southern_and_Southeastern_Asia x 2018	1.001^{***} (0.013)	1.472^{***} (0.059)
FSC x Southern_and_Southeastern_Asia x 2019	1.031^{***} (0.016)	3.344^{***} (0.513)
Constant	-1.554^{***} (0.026)	7.999^{***} (0.913)
Observations	34,794	34,794
Adjusted R^2	0.649	-1.006
Residual Std. Error (df = 34733)	0.484	1.156
F Statistic	$1,072.535^{***}$ (df = 60; 34733)	
Wu-Hausman stat.		723.996^{***} (df=2,34731)
Weak IV stat. (Expected relative rigor)		343.283^{***} (df=2,34733)
Weak IV stat. (Price at certifier-market-level in 1K USD)		77.239^{***} (df=2,34733)

Notes: Heteroscedasticity robust standard errors. Standard errors are not yet corrected for the use of generated regressors and regressands. The dependent variable are the estimates of mean utility by certifier and market, obtained from Maximum Likelihood estimation of the certifier-market constants in the nested logit model of FMUs' certifier choice. The outside option in the nested logit includes both dropping out of FSC certification or not getting certified yet. Dropping out is permanent in the data, such that only the first year after drop-out is included. In the 2SLS regression here, the outside option is excluded since its mean utility is normalized to zero for all markets in the nested logit.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A8: First stage of IV: Price prediction

	(1)	(2)
β^x:		
First year certifier is available	-0.154^{***} (0.024)	-0.131^{***} (0.025)
Average distance to certifier's headquarter by market in 1000 km	0.148^{***} (0.003)	0.147^{***} (0.003)
Excluded instruments:		
Close-by rivals in terms of certifier's experience		-0.119^{***} (0.006)
Close-by rivals in terms of certifier's experience x Forest country's longitude (mean by market)		-0.000^{**} (0.000)
ξ:		
Certifier 2	0.222^{***} (0.010)	0.345^{***} (0.012)
Certifier 3	-0.866^{***} (0.012)	-0.729^{***} (0.014)
Certifier 4	0.905^{***} (0.011)	1.212^{***} (0.019)
Certifier 5	-0.137^{***} (0.017)	0.004 (0.018)
Certifier 6	1.085^{***} (0.024)	1.209^{***} (0.023)
Certifier 7	0.192^{***} (0.016)	0.347^{***} (0.019)
Small Certifiers	0.532^{***} (0.011)	0.791^{***} (0.020)
FSC x Brazil x 2015	-0.461^{***} (0.045)	-0.614^{***} (0.046)
FSC x Brazil x 2016	-0.584^{***} (0.045)	-0.673^{***} (0.045)
FSC x Brazil x 2017	-0.619^{***} (0.045)	-0.703^{***} (0.046)
FSC x Brazil x 2018	-0.774^{***} (0.045)	-0.858^{***} (0.046)
FSC x Brazil x 2019	-0.304^{***} (0.045)	-0.388^{***} (0.046)
FSC x Central_America x 2015	-1.112^{***} (0.044)	-1.324^{***} (0.046)
FSC x Central_America x 2016	-0.850^{***} (0.045)	-1.062^{***} (0.047)
FSC x Central_America x 2017	-0.959^{***} (0.047)	-1.172^{***} (0.049)
FSC x Central_America x 2018	-1.308^{***} (0.044)	-1.521^{***} (0.046)
FSC x Central_America x 2019	-0.760^{***} (0.048)	-0.973^{***} (0.050)
FSC x Eastern_Asia_excl_China x 2015	-0.181^{*} (0.094)	-0.298^{***} (0.094)
FSC x Eastern_Asia_excl_China x 2016	-0.186^{**} (0.095)	-0.302^{***} (0.094)
FSC x Eastern_Asia_excl_China x 2017	-0.164^{*} (0.093)	-0.281^{***} (0.093)
FSC x Eastern_Asia_excl_China x 2018	-0.153 (0.095)	-0.270^{***} (0.095)
FSC x Eastern_Asia_excl_China x 2019	0.398^{***} (0.129)	0.275^{**} (0.128)
FSC x Eastern_Europe_excl_CIS x 2015	0.098^{*} (0.052)	-0.041 (0.052)
FSC x Eastern_Europe_excl_CIS x 2016	0.136^{***} (0.052)	-0.002 (0.053)
FSC x Eastern_Europe_excl_CIS x 2017	0.120^{**} (0.052)	-0.018 (0.053)
FSC x Eastern_Europe_excl_CIS x 2018	-0.091^{*} (0.048)	-0.148^{***} (0.049)
FSC x Eastern_Europe_excl_CIS x 2019	0.346^{***} (0.053)	0.196^{***} (0.055)
FSC x Eastern_and_Southern_Africa x 2015	-0.565^{***} (0.097)	-0.819^{***} (0.097)
FSC x Eastern_and_Southern_Africa x 2016	-0.649^{***} (0.097)	-0.895^{***} (0.097)
FSC x Eastern_and_Southern_Africa x 2017	-0.737^{***} (0.095)	-0.983^{***} (0.095)
FSC x Eastern_and_Southern_Africa x 2018	-0.870^{***} (0.096)	-1.116^{***} (0.096)
FSC x Eastern_and_Southern_Africa x 2019	-0.483^{***} (0.102)	-0.729^{***} (0.101)
FSC x Northern_America x 2015	-0.235^{***} (0.045)	-0.307^{***} (0.046)
FSC x Northern_America x 2016	-0.252^{***} (0.045)	-0.324^{***} (0.046)
FSC x Northern_America x 2017	-0.203^{***} (0.045)	-0.276^{***} (0.046)
FSC x Northern_America x 2018	-0.824^{***} (0.043)	-0.897^{***} (0.044)
FSC x Northern_America x 2019	0.114^{**} (0.047)	0.042 (0.048)
FSC x Oceania x 2015	1.538^{***} (0.106)	1.425^{***} (0.108)
FSC x Oceania x 2016	-0.242^{***} (0.075)	-0.355^{***} (0.077)
FSC x Oceania x 2017	-0.097 (0.076)	-0.210^{***} (0.078)
FSC x Oceania x 2018	-0.307^{***} (0.073)	-0.419^{***} (0.075)
FSC x Oceania x 2019	0.761^{***} (0.096)	0.648^{***} (0.098)
FSC x South_America_excl_Brazil x 2015	-0.933^{***} (0.049)	-0.952^{***} (0.051)
FSC x South_America_excl_Brazil x 2016	-1.023^{***} (0.049)	-1.043^{***} (0.050)
FSC x South_America_excl_Brazil x 2017	-1.115^{***} (0.048)	-1.134^{***} (0.050)
FSC x South_America_excl_Brazil x 2018	-1.081^{***} (0.050)	-1.100^{***} (0.051)
FSC x South_America_excl_Brazil x 2019	-0.606^{***} (0.051)	-0.625^{***} (0.052)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2016	-0.001 (0.049)	-0.001 (0.049)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2017	0.019 (0.047)	0.133^{***} (0.047)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2018	-0.205^{***} (0.046)	-0.088^{*} (0.046)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2019	0.248^{***} (0.050)	0.296^{***} (0.049)
FSC x Southern_and_Southeastern_Asia x 2015	0.284^{***} (0.049)	0.333^{***} (0.052)
FSC x Southern_and_Southeastern_Asia x 2016	0.627^{***} (0.049)	0.676^{***} (0.053)
FSC x Southern_and_Southeastern_Asia x 2017	0.005 (0.048)	0.055 (0.051)
FSC x Southern_and_Southeastern_Asia x 2018	0.112^{**} (0.049)	0.162^{***} (0.052)
FSC x Southern_and_Southeastern_Asia x 2019	2.745^{***} (0.062)	2.794^{***} (0.066)
Constant	2.173^{***} (0.035)	2.410^{***} (0.038)
Observations	34,794	34,794
Adjusted R^2	0.610	0.611
Residual Std. Error	0.833 (df = 34735)	0.831 (df = 34733)
F Statistic	937.949^{***} (df = 58; 34735)	913.239^{***} (df = 60; 34733)

Notes: First stage of 2SLS regression of estimates of mean utility by certifier and market (obtained from MLE of the certifier-market constants in the nested logit demand model) on prices and other characteristics. The dependent variable in the first stage is the endogeneous certifier-market level prediction of prices. Standard errors are heteroscedasticity and serial correlation robust, but not yet corrected for the use of generated regressors and regressands.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A9: First stage of IV: Prediction of expected relative rigor

	(1)	(2)
β^x:		
First year certifier is available	0.094*** (0.010)	0.095*** (0.009)
Average distance to certifier's headquarter by market in 1000 km	−0.014*** (0.000)	−0.013*** (0.000)
Excluded instruments:		
Close-by rivals in terms of certifier's experience		0.013*** (0.001)
Close-by rivals in terms of certifier's experience x Forest country's longitude (mean by market)		−0.000*** (0.000)
ξ:		
Certifier 2	−0.148*** (0.001)	−0.163*** (0.001)
Certifier 3	0.113*** (0.001)	0.095*** (0.001)
Certifier 4	−0.028*** (0.001)	−0.062*** (0.002)
Certifier 5	−0.171*** (0.001)	−0.188*** (0.001)
Certifier 6	0.204*** (0.001)	0.188*** (0.001)
Certifier 7	−0.159*** (0.001)	−0.178*** (0.001)
Small Certifiers	0.040*** (0.002)	0.004* (0.002)
FSC x Brazil x 2015	−0.013*** (0.002)	−0.002 (0.002)
FSC x Brazil x 2016	−0.010*** (0.003)	−0.008*** (0.003)
FSC x Brazil x 2017	−0.019*** (0.002)	−0.018*** (0.002)
FSC x Brazil x 2018	0.002 (0.002)	0.003 (0.002)
FSC x Brazil x 2019	0.024*** (0.003)	0.026*** (0.003)
FSC x Central_America x 2015	0.010*** (0.002)	0.023*** (0.002)
FSC x Central_America x 2016	−0.000 (0.002)	0.012*** (0.002)
FSC x Central_America x 2017	0.006*** (0.002)	0.019*** (0.002)
FSC x Central_America x 2018	−0.001 (0.002)	0.011*** (0.002)
FSC x Central_America x 2019	−0.001 (0.002)	0.011*** (0.002)
FSC x Eastern_Asia_excl_China x 2015	−0.010*** (0.003)	0.019*** (0.003)
FSC x Eastern_Asia_excl_China x 2016	−0.008*** (0.003)	0.021*** (0.003)
FSC x Eastern_Asia_excl_China x 2017	−0.010*** (0.003)	0.019*** (0.003)
FSC x Eastern_Asia_excl_China x 2018	0.079*** (0.012)	0.108*** (0.012)
FSC x Eastern_Asia_excl_China x 2019	−0.001 (0.003)	0.027*** (0.003)
FSC x Eastern_Europe_excl_CIS x 2015	0.093*** (0.009)	0.111*** (0.009)
FSC x Eastern_Europe_excl_CIS x 2016	−0.008*** (0.002)	0.010*** (0.002)
FSC x Eastern_Europe_excl_CIS x 2017	0.004** (0.002)	0.022*** (0.002)
FSC x Eastern_Europe_excl_CIS x 2018	−0.025*** (0.003)	−0.015*** (0.003)
FSC x Eastern_Europe_excl_CIS x 2019	0.003 (0.002)	0.022*** (0.002)
FSC x Eastern_and_Southern_Africa x 2015	−0.055*** (0.005)	−0.027*** (0.005)
FSC x Eastern_and_Southern_Africa x 2016	−0.024*** (0.003)	0.004 (0.003)
FSC x Eastern_and_Southern_Africa x 2017	−0.018*** (0.004)	0.010*** (0.004)
FSC x Eastern_and_Southern_Africa x 2018	0.003 (0.002)	0.031*** (0.003)
FSC x Eastern_and_Southern_Africa x 2019	−0.019*** (0.003)	0.009** (0.004)
FSC x Northern_America x 2015	0.013*** (0.002)	0.006*** (0.002)
FSC x Northern_America x 2016	0.006** (0.002)	−0.001 (0.002)
FSC x Northern_America x 2017	0.005** (0.002)	−0.002 (0.002)
FSC x Northern_America x 2018	0.020*** (0.002)	0.013*** (0.002)
FSC x Northern_America x 2019	0.011*** (0.002)	0.004** (0.002)
FSC x Oceania x 2015	0.029*** (0.004)	0.060*** (0.004)
FSC x Oceania x 2016	0.032*** (0.003)	0.062*** (0.004)
FSC x Oceania x 2017	0.044*** (0.004)	0.074*** (0.004)
FSC x Oceania x 2018	0.028*** (0.005)	0.058*** (0.005)
FSC x Oceania x 2019	0.020*** (0.003)	0.050*** (0.003)
FSC x South_America_excl_Brazil x 2015	−0.006*** (0.002)	−0.020*** (0.002)
FSC x South_America_excl_Brazil x 2016	0.059*** (0.004)	0.046*** (0.004)
FSC x South_America_excl_Brazil x 2017	0.016*** (0.003)	0.003 (0.003)
FSC x South_America_excl_Brazil x 2018	−0.001 (0.002)	−0.015*** (0.002)
FSC x South_America_excl_Brazil x 2019	0.002 (0.002)	−0.012*** (0.002)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2016	−0.018*** (0.002)	−0.018*** (0.002)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2017	0.087*** (0.005)	0.075*** (0.005)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2018	0.015*** (0.003)	0.003 (0.003)
FSC x Southern_Europe_incl_Turkey_excl_Italy x 2019	0.024*** (0.004)	0.020*** (0.003)
FSC x Southern_and_Southeastern_Asia x 2015	0.008*** (0.002)	0.026*** (0.003)
FSC x Southern_and_Southeastern_Asia x 2016	0.000 (0.002)	0.018*** (0.003)
FSC x Southern_and_Southeastern_Asia x 2017	0.000 (0.002)	0.018*** (0.003)
FSC x Southern_and_Southeastern_Asia x 2018	0.025*** (0.002)	0.042*** (0.003)
FSC x Southern_and_Southeastern_Asia x 2019	0.024*** (0.003)	0.042*** (0.003)
Constant	0.502*** (0.002)	0.477*** (0.002)
Observations	34,794	34,794
Adjusted R^2	0.884	0.886
Residual Std. Error	0.056 (df = 34735)	0.055 (df = 34733)
F Statistic	4,553.911*** (df = 58; 34735)	4,500.315*** (df = 60; 34733)

Notes: First stage of 2SLS regression of estimates of mean utility by certifier and market (obtained from MLE of the certifier-market constants in the nested logit demand model) on prices and other characteristics. The dependent variable in the first stage is the endogeneous certifier-market level prediction of expected relative rigor. Standard errors are heteroscedasticity and serial correlation robust, but not yet corrected for the use of generated regressors and regressands.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table A10: Estimates of heterogeneity in preferences across FMU types (Nested Logit)

$\tilde{\alpha}$: Price at certifier-market-level in 1K USD X ...	
Cost factor	0.020*** (0.008)
$\tilde{\boldsymbol{\beta}}^{\boldsymbol{r}}$: Expected relative rigor X ...	
Expected baseline violations	0.082 (0.097)
$\geq$ 5 violations reports expected from most rigorous certifier	0.519 (1.290)
Forest chose same certifier last year	2.496*** (0.345)
Forest chose same certifier last year x Yrs. with FSC cert.	0.051*** (0.014)
Log of forest's country's score in Corruption Perceptions Index x FSC	0.399 (0.316)
Forest has angiosperms (clade of plants) x FSC	0.167 (0.190)
Forest has gymnosperms (clade of plants) x FSC	0.513*** (0.145)
Forest has myrtaceae (clade of plants) x FSC	0.141 (0.147)
Country's export value of wood chips in 1000 USD per cubic metre	−0.273*** (0.097)
Certifier's experience in forest's first certification year	0.046*** (0.017)
Forest chose same certifier last year x Recertification year x FSC	−0.076 (0.166)
FSC Certifier has office in forest's country	0.549*** (0.141)
Is first FSC certifier (entry cost)	−2.457*** (0.281)
$\boldsymbol{\xi}$:	
Group certificate (vs. individual) X ...	
Certifier 1	−0.046 (0.149)
Certifier 2	−0.373 (0.328)
Certifier 3	−0.098 (0.287)
Certifier 4	0.205 (0.254)
Certifier 5	−0.004 (0.206)
Certifier 6	0.135 (0.191)
Certifier 7	−0.309 (0.280)
Small Certifiers	−0.598** (0.289)
Forest is classified as plantation X ...	
Certifier 1	−0.568*** (0.193)
Certifier 2	0.801*** (0.300)
Certifier 3	−0.220 (0.351)
Certifier 4	0.631*** (0.232)
Certifier 5	−0.481* (0.265)
Certifier 6	0.345 (0.212)
Certifier 7	−0.009 (0.254)
Small Certifiers	−0.317 (0.276)
λ:	
Within FSC nest correlation	0.464*** (0.063)
Certifier-market FE	Yes
Observations (choice situations)	6,250
Log Likelihood	−2,744.536

Notes: Maximum Likelihood estimation of nested logit choice model with R package mlogit. Heteroscedasticity and serial correlation robust standard errors. Standard errors are not yet corrected for the use of generated regressors. The outside option includes both dropping out of FSC certification or not getting certified yet. Dropping out is permanent in the data, such that only the first year after drop-out is included. The mean utility of the outside option is normalized to zero for all markets.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

C.2.3 Robustness and tests

I test the relevance of the differentiation instruments used in the demand model in two ways. First, Table 7 shows the statistics and significance levels of tests for weak instruments, as provided by the *ivreg* package (Zeileis et al., 2023). For both prices and expected relative rigor, these tests allow me to reject the null hypothesis that the differentiation instruments are jointly insignificant in the first stage, i.e., in predicting the endogenous variable conditional on all other controls. Second, I follow Stock and Yogo (2005) to test the relevance of the instruments in a way that is adapted to the case of multiple endogenous variables. That is, I test for the joint significance of the differentiation instruments in both first stages jointly, using the adapted critical values derived by Stock and Yogo (2005). The corresponding F test statistic is 177.577, which is much larger than 8.2, the critical value for 2 endogenous variables and 2 instruments (Hansen, 2021). I, thus, reject the null hypothesis of weak instruments. A caveat is that both tests assume homoskedasticity. However, to my knowledge, no adapted heteroskedasticity-robust test zxists for weak instruments in the case of multiple endogenous variables (Andrews, 2018).

Table A11: Preference estimates at the certifier-market-levelwith region and year instead of market (region-year) fixed effects

	OLS	*2SLS*
	(1)	(2)
$\bar{\alpha}$: Price at certifier-market-level in 1K USD	-0.077^{***} (0.004)	-1.098^{***} (0.195)
$\bar{\beta}^r$: Expected relative rigor	-1.319^{***} (0.054)	-13.160^{***} (0.748)
$\boldsymbol{\beta^x}$:		
First year certifier is available	0.448^{***} (0.019)	1.659^{***} (0.171)
Average distance to certifier's headquarter by market in 1000 km	-0.024^{***} (0.001)	-0.126^{***} (0.010)
$\boldsymbol{\xi}$:		
Certifier 2	-1.235^{***} (0.017)	-2.540^{***} (0.086)
Certifier 3	-0.490^{***} (0.014)	-0.052 (0.131)
Certifier 4	-0.786^{***} (0.013)	-0.317^{**} (0.148)
Certifier 5	-0.931^{***} (0.015)	-3.105^{***} (0.145)
Certifier 6	-0.615^{***} (0.015)	2.818^{***} (0.327)
Certifier 7	-1.025^{***} (0.014)	-2.417^{***} (0.091)
Certifier Small Certifiers	-0.321^{***} (0.018)	1.007^{***} (0.192)
Central_America	0.176^{***} (0.014)	0.058 (0.035)
Eastern_and_Southern_Africa	0.205^{***} (0.025)	-0.496^{***} (0.105)
Eastern_Asia_excl_China	0.114^{***} (0.028)	0.413^{***} (0.079)
Eastern_Europe_excl_CIS	-0.005 (0.017)	0.280^{***} (0.054)
Northern_America	-0.796^{***} (0.014)	-1.198^{***} (0.100)
Oceania	0.096^{***} (0.024)	1.682^{***} (0.266)
South_America_excl_Brazil	-0.076^{***} (0.012)	-0.134^{***} (0.045)
Southern_and_Southeastern_Asia	0.334^{***} (0.012)	1.296^{***} (0.161)
Southern_Europe_incl_Turkey_excl_Italy	-0.402^{***} (0.015)	0.230^{**} (0.093)
2016	0.022^{**} (0.010)	-0.002 (0.020)
2017	-0.262^{***} (0.010)	-0.231^{***} (0.022)
2018	0.402^{***} (0.009)	0.265^{***} (0.036)
2019	0.235^{***} (0.011)	1.081^{***} (0.158)
Constant	-1.375^{***} (0.035)	6.705^{***} (0.710)
Observations	34,794	34,794
Adjusted R^2	0.484	-1.482
Residual Std. Error (df = 34769)	0.575	1.262
F Statistic	$1{,}359.602^{***}$ (df = 24; 34769)	
Wu-Hausman stat.		454.37^{***} (df=2,34767)
Weak IV stat. (Expected relative rigor)		396.568^{***} (df=2,34769)
Weak IV stat. (Price at certifier-market-level in 1K USD)		49.082^{***} (df=2,34769)

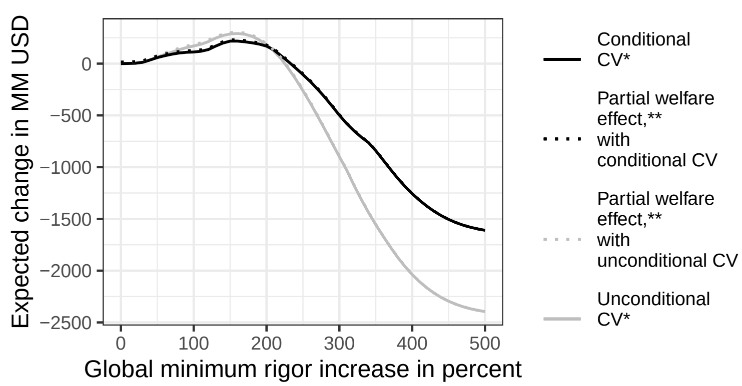
Notes: Heteroscedasticity robust standard errors. Standard errors are not yet corrected for the use of generated regressors and regressands. The dependent variable are the estimates of mean utility by certifier and market, obtained from Maximum Likelihood estimation of the certifier-market constants in the nested logit model of FMUs' certifier choice. The outside option in the nested logit includes both dropping out of FSC certification or not getting certified yet. Dropping out is permanent in the data, such that only the first year after drop-out is included. In the 2SLS regression here, the outside option is excluded since its mean utility is normalized to zero for all markets in the nested logit.$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

D Appendix: Counterfactual Analysis

D.1 Robustness of the measure of consumer valuation

In the counterfactual analysis, I assume that consumer valuation is a linear function of the amount of certified wood and the weighted sum of violation reports conditional on certification. This assumption can be debated in two dimensions. First, there is no empirical evidence to microfound the functional form. The marginal benefits for consumers and society of additional violation reports may as well decrease for higher quality, such that the plotted welfare estimates would overestimate true benefits. Second, the assumption implies that consumer valuation depends on the amount of certified wood and certification quality. This seems to be the most plausible assumption since this type of conditional valuation relates more to the traditional way of comparing products of varying quality. Nevertheless, other heuristics of valuing certification are possible.

Figure A21: Expected changes following direct shifts in minimum rigor



Notes: *Consumer valuation of FSC certification, assuming valuation at 10 percent of wood prices for 2017 baseline quality and constant wood prices. **Sum of CV, FMU and certifier surplus if price premia at baseline CV. Excl. externalities not valued by consumers.

In Figure A21, I consider the possibility of consumer valuation varying with the quality among all once-certified FMUs, i.e., the total impact of FSC in the industry, *vsum*, but not directly with the amount of certified wood. In other words, consumer valuation might work more like donations rather than price premia. Figure A21 shows changes in such unconditional consumer valuation and corresponding welfare effects in gray. For comparison, the black lines show the results using conditional consumer valuation presented in the paper. The effects are qualitatively similar. Unconditional consumer valuation seems to amplify the changes suggested by conditional consumer valuation mostly.

D.2 Back-of-the-envelope calculation of welfare benefit of reduced tree cover loss through FSC violation reports

An estimate of FSC's effect on deforestation allows me to obtain an idea of the benefits from corresponding violations: for the Congo Basin, Tritsch et al. (2020) estimate an average reduction in deforestation by 514 ha per FMU from 2000 to 2010 due to FSC certification. CO2 emissions from deforestation vary largely across contexts. If the difference in CO2 emissions between forested and deforested areas in the Congo Basin is similar to that in Brazil, I can use the average from there, 80 tCO2/ha (Souza-Rodrigues, 2019). Using a carbon cost of 31 USD/tCO2 (Nordhaus, 2017), the total social benefit of FSC certification per year and FMU would be 127,472 USD. Using the average number of violation reports per FMU and the fraction of those relating to tree cover loss in a randomly drawn sample, one violation report's carbon benefit is estimated at roughly 2 million USD, while this paper's measure of consumer valuation per violation report is 2.79 million USD. The derived estimate of carbon benefits should be interpreted cautiously, as the calculation is based on a mix of different sources.

D.3 Suspending the accreditation of lenient certifiers - The role of price changes

Table A12: Expected changes in violation reports

Counterfactual	Conditional on certification in numbers	Due to drop-outs in numbers	Total in numbers	Total in percent
(1) Targeting the most lenient certifier				
Accreditation suspension				
Effect through				
- Certifier choices and prices	62.36	-47.58	14.77	2.51
- Certifier choices only	63.90	-46.06	17.83	3.03
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	18.62	-14.34	4.28	0.73
- Certifier choices only	19.02	-14.45	4.56	0.78
(2) Targeting the two most lenient certifiers				
Accreditation suspension				
Effect through				
- Certifier choices and prices	120.59	-98.68	21.91	3.72
- Certifier choices only	122.68	-92.14	30.54	5.19
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	64.65	-45.26	19.39	3.29
- Certifier choices only	65.23	-44.33	20.89	3.55
(3) Targeting the three most lenient certifiers				
Accreditation suspension				
Effect through				
- Certifier choices and prices	145.39	-116.74	28.65	4.87
- Certifier choices only	146.18	-106.97	39.21	6.66
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	133.10	-86.56	46.54	7.90
- Certifier choices only	134.63	-80.55	54.08	9.19

Notes: Changes in counterfactual scenarios compared to baseline. Accreditation withdrawal is implemented by removing the corresponding certifier from the choiceset. Direct shift of minimum rigor is implemented by shifting the targeted certifiers to the next most lenient certifier's rigor in each market.

Table A13: Expected change in surplus from FSC certification

Counterfactual	FMUs	Targeted certifiers	Untargeted certifiers	Consumers	Total	
	in MM USD	in MM USD	in MM USD	in MM USD	in MM USD	in percent
(1) Targeting the most lenient certifier						
Accreditation suspension						
Effect through						
- Certifier choices and prices	-0.51	-1.53	1.20	23.91	23.06	150.96
- Certifier choices only	-0.49	-1.53	1.18	31.76	30.92	202.42
Equivalent minimum rigor shift						
Effect through						
- Certifier choices and prices	-0.10	-0.21	0.11	16.01	15.81	103.51
- Certifier choices only	-0.10	-0.21	0.11	16.34	16.14	105.69
(2) Targeting the two most lenient certifiers						
Accreditation suspension						
Effect through						
- Certifier choices and prices	-0.92	-2.82	1.88	16.42	14.56	95.33
- Certifier choices only	-0.84	-2.82	1.83	45.40	43.58	285.27
Equivalent minimum rigor shift						
Effect through						
- Certifier choices and prices	-0.47	-1.08	0.47	26.94	25.86	169.27
- Certifier choices only	-0.45	-1.11	0.49	31.09	30.02	196.55
(3) Targeting the three most lenient certifiers						
Accreditation suspension						
Effect through						
- Certifier choices and prices	-1.07	-3.48	2.36	7.96	5.77	37.76
- Certifier choices only	-0.97	-3.48	2.28	42.28	40.11	262.58
Equivalent minimum rigor shift						
Effect through						
- Certifier choices and prices	-0.87	-2.38	1.46	49.14	47.34	309.92
- Certifier choices only	-0.80	-2.44	1.44	75.32	73.51	481.24

Notes: Changes in counterfactual scenarios compared to baseline. Accreditation withdrawal is implemented by removing the corresponding certifier from the choiceset. Direct shift of minimum rigor is implemented by shifting the targeted certifiers to the next most lenient certifier's rigor in each market.

Table A14: Simulated changes in audit quality and participation

Counterfactual	Avg. minimum rigor across markets (mechanical change)		Participating FMUs (expected change)	
	in numbers	in percent	in numbers	in percent
(1) Targeting the most lenient certifier				
Accreditation suspension				
Effect through				
- Certifier choices and prices	0.04	9.73	-57.16	-5.92
- Certifier choices only	0.04	9.73	-53.47	-5.54
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	0.04	9.73	-11.13	-1.15
- Certifier choices only	0.04	9.73	-11.24	-1.16
(2) Targeting the two most lenient certifiers				
Accreditation suspension				
Effect through				
- Certifier choices and prices	0.08	20.51	-124.12	-12.85
- Certifier choices only	0.08	20.51	-109.26	-11.31
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	0.08	20.51	-57.21	-5.92
- Certifier choices only	0.08	20.51	-54.60	-5.65
(3) Targeting the three most lenient certifiers				
Accreditation suspension				
Effect through				
- Certifier choices and prices	0.11	29.25	-145.86	-15.10
- Certifier choices only	0.11	29.25	-127.62	-13.22
Equivalent minimum rigor shift				
Effect through				
- Certifier choices and prices	0.11	29.25	-103.32	-10.70
- Certifier choices only	0.11	29.25	-92.67	-9.60

Notes: Changes in counterfactual scenarios compared to baseline. Accreditation withdrawal is implemented by removing the corresponding certifier from the choiceset. Direct shift of minimum rigor is implemented by shifting the targeted certifiers to the next most lenient certifier's rigor in each market.

References

Alecci, S. (2023). Environmental auditors approve green labels for products linked to deforestation and authoritarian regimes. https://www.icij.org/investigations/deforestation-inc/auditors-green-labels-sustainability-environmental-harm/. Last visited 2023-10-03.

Alix-Garcia, J. and Wolff, H. (2014). Payment for Ecosystem Services from Forests. *Annual Review of Resource Economics*, 6:361–380.

Andrews, I. (2018). Weak Instruments and What to Do about Them. https://www.nber.org/lecture/2018-methods-lecture-isaiah-andrews-weak-instruments-and-what-do-about-them. Last visited 2023-08-02.

Araujo, M., Kant, S., and Couto, L. (2009). Why Brazilian companies are certifying their forests? *Forest Policy and Economics*, 11(8):579–585.

ASI (2019). Procedure on Surveillance & Sampling - ASI-PRO-20-105-Surveillance & Sampling-V6.4. https://asi-login.my.salesforce.com/sfc/p/#A0000000aGza/a/1H000000HqeR/i9Ry6Lmd8ybwkdMnw1BNjQU0TbSKoq2tObMHfOHrwgA. Last visited 2022-04-09.

ASI (2020). Find a CAB. https://www.asi-assurance.org/s/find-a-cab. Last visited 2021-09-23.

ASI (2021). Procedure for Witness and Compliance Assessments - ASI-PRO-20-111-Witness & Compliance Assessments-V2.3. https://asi-login.my.salesforce.com/sfc/p/#A0000000aGza/a/120000000GKD/fJyFDEdbdnpj_qHotX7mJRDcjYsq1cbW2jXHgVGmqPE. Last visited 2022-09-04.

ASI (2022). Assurance Services International. https://www.asi-assurance.org. Last visited 2022-08-02.

Assunção, J., McMillan, R., Murphy, J., and Souza-Rodrigues, E. (2023). Optimal environmental targeting in the Amazon rainforest. *The Review of Economic Studies*, 90(4):1608–1641.

Auriol, E. and Schilizzi, S. G. (2015). Quality signaling through certification in developing countries. *Journal of Development Economics*, 116:105–121.

Balboni, C., Burgess, R., and Olken, B. A. (2021). The Origins and Control of Forest Fires in the Tropics. Working Paper.

Bar, T. and Zheng, Y. (2019). Choosing certifiers: Evidence from the British Retail Consortium food safety standard. *American Journal of Agricultural Economics*, 101(1):74–88.

Becker, B. and Milbourn, T. (2011). How did increased competition affect credit ratings? *Journal of Financial Economics*, 101(3):493–514.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, 63(4):841–890.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 25(2):242–262.

Bizzotto, J. and Harstad, B. (2020). The Choice of Certifier in Endogenous Markets. *SSRN Electronic Journal*.

Blackman, A., Goff, L., and Rivera Planter, M. (2018). Does eco-certification stem tropical deforestation? Forest Stewardship Council certification in Mexico. *Journal of Environmental Economics and Management*, 89:306–333.

Blackman, A., Raimondi, A., and Cubbage, F. (2017). Does forest certification in developing countries have environmental benefits? Insights from Mexican corrective action requests. *International Forestry Review*, 19(3):247–264.

Blackman, A. and Rivera, J. (2011). Producer-Level Benefits of Sustainability Certification. *Conservation Biology*, 25(6):1176–1185.

Bolton, P., Freixas, X., and Shapiro, J. (2012). The credit ratings game. *The Journal of Finance*, 67(1):85–111.

Briesch, R. A., Dillon, W. R., and Blattberg, R. C. (2008). Treating Zero Brand Sales Observations in Choice Model Estimation: Consequences and Potential Remedies. *Journal of Marketing Research (JMR)*, 45(5):618–632.

Burgess, R., Hansen, M., Olken, B. A., and Potapov, P. (2012). The political economy of deforestation in the tropics. *The Quarterly Journal of Economics*, 127(4):1707–1754.

Cai, Z. and Aguilar, F. X. (2013). Meta-analysis of consumer's willingness-to-pay premiums for certified wood products. *Journal of Forest Economics*, 19(1):15–31.

Cameron, A. C. and Windmeijer, F. (1996). R-Squared Measures for Count Data Regression Models with Applications to Health-Care Utilization. *Journal of Business & Economic Statistics*, 14(2):209–220.

Chen, J., Innes, J., and Tikina, A. (2010). Private cost-benefits of voluntary forest product certification. *International Forestry Review*, 12(1):1–12.

Chu, J., Fisman, R., Tan, S., and Wang, Y. (2021). Hometown Ties and the Quality of Government Monitoring: Evidence from Rotation of Chinese Auditors. *American Economic Journal: Applied Economics*, 13(3):176–201.

Clark, M. and Kozar, J. (2011). Comparing Sustainable Forest Management Certifications Standards: A Meta-analysis. *Ecology and Society*, 16(1).

Conniff, R. (2018). Greenwashed Timber: How Sustainable Forest Certification Has Failed. https://e360.yale.edu/features/greenwashed-timber-how-sustainable-forest-certification-has-failed. Last visited 2023-07-03.

Cubbage, F., Moore, S., Henderson, T., and Araujo, M. M. F. C. (2009). Costs and benefits of forest certification in the Americas. In *Natural Resources: Management, Economic Development and Protection*, pages 155–183. Nova Publishers.

Darby, M. R. and Karni, E. (1973). Free competition and the optimal amount of fraud. *The Journal of Law and Economics*, 16(1):67–88.

DeFries, R. S., Fanzo, J., Mondal, P., Remans, R., and Wood, S. A. (2017). Is voluntary certification of tropical agricultural commodities achieving sustainability goals for small-scale producers? A review of the evidence. *Environmental Research Letters*, 12(3):033001.

D'Haultfoeuille, X., Durrmeyer, I., and Février, P. (2019). Automobile prices in market equilibrium with unobserved price discrimination. *The Review of Economic Studies*, 86(5):1973–1998.

Dragusanu, R., Giovannucci, D., and Nunn, N. (2014). The Economics of Fair Trade. *Journal of Economic Perspectives*, 28(3):217–236.

Dranove, D. and Jin, G. Z. (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature*, 48(4):935–63.

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

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2013b). What does reputation buy? Differentiation in a market for third-party auditors. *American Economic Review: Papers and Proceedings*, 103(3):314–19.

Duflo, E., Greenstone, M., Pande, R., and Ryan, N. (2018). The Value of Regulatory Discretion: Estimates From Environmental Inspections in India. *Econometrica*, 86(6):2123–2160.

Earthsight (2020). FLATPÅCKED FÖRESTS. https://www.earthsight.org.uk/flatpackedforests-en#group-Summary-rayWKL9K1I. Last visited 2022-01-31.

Earthsight (2021). Open letter: FSC is no longer fit for purpose and must urgently reform. https://www.earthsight.org.uk/news/blog-open-letter-fsc-no-longer-fit-for-purpose-and-must-urgently-reform. Last visited 2022-07-18.

Egger, P., Larch, M., Staub, K. E., and Winkelmann, R. (2011). The Trade Effects of Endogenous Preferential Trade Agreements. *American Economic Journal: Economic Policy*, 3:113–143.

European Commission (2018). EU Law on Organic Production: An Overview. https://ec.europa.eu/agriculture/organic/eu-policy/eu-legislation/brief-overview_en. Last visited 2021-09-23.

European Union (2023). REGULATION (EU) 2023/1115 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation and repealing Regulation (EU) No 995/2010. *Official Journal of the European Union*, L 150:206–247.

FAO (2018). State of the world's forests 2018 - Forests pathways to sustainable development. Technical report, FAO, Rome.

FAO (2020). Global Forest Resources Assessment 2020: Main report. Technical report, FAO, Rome.

FAO (2023). Forestry Production and Trade. https://www.fao.org/faostat/en/#data/FO. Last visited 2023-04-15.

Federal Ministry for Economic Cooperation and Development (2023). Green Button | Grüner Knopf. https://www.gruener-knopf.de/en. Last visited 2023-08-12.

FRED (2020). Producer Price Index by Commodity: Pulp, Paper, and Allied Products: Wood Pulp. https://fred.stlouisfed.org/series/WPU0911. Last visited 2021-09-23.

Frey, G., Cubbage, F., Ha, T., Davis, R., Carle, J., Thon, V., and Dzung, N. (2018). Financial analysis and comparison of smallholder forest and state forest enterprise plantations in Central Vietnam. *International Forestry Review*, 20(2):181–198.

FSC (2004). SLIMF Eligibility Criteria (FSC-STD-01-003 (Version 1-0) EN). https://fsc.org/sites/default/files/2021-07/FSC-STD-01-003%20SLIMF%20Eligibility%20Criteria.pdf. Last visited 2021-09-23.

FSC (2009a). Forest management evaluations addendum – Forest certification public summary reports (FSC-STD-20-007b (V1-0) EN). https://connect.fsc.org/document-centre/documents/resource/253. Last visited 2023-10-03.

FSC (2009b). Forest management evaluations (FSC-STD-20-007 (V3-0) EN). https://connect.fsc.org/document-center/documents/0bbbebcf-81af-4e89-9c88-3d6e4ed6581d. Last visited 2021-10-13.

FSC (2010). Transfer of FSC Certificates and License Agreements (FSC-PRO-20-003 (V1-0) EN). https://connect.fsc.org/document-centre/documents/resource/236. Last visited 2023-05-17.

FSC (2015). FSC principles and criteria for Forest Stewardship (FSC-STD-01-001 V5-2 EN). https://connect.fsc.org/current-processes/fsc-std-01-001-v5-2-fsc-principles-and-criteria-pc-forest-stewardship. Last visited 2021-10-15.

FSC (2015). General requirements for FSC accredited certification bodies (FSC-STD-20-001 V4-0 EN). https://connect.fsc.org/document-centre/documents/resource/280. Last visited 2021-10-13.

FSC (2016a). Annual Administration Fee (AAF) (FSC-POL-20-005 V2-5 EN). https://fsc.org/en/document-centre. Last visited 2020-04-26.

FSC (2016b). Cameroon Recognizes FSC Forest Management Certification as FLEGT-Compliant: A Welcome Precedent. https://fsc.org/sites/default/files/2019-04/

Cameroon%20first%20country%20to%20recognise%20FSC%20FLEGT%20compliant%20%281%29.pdf.

FSC (2017). FSC Glossary of Terms (FSC-STD-01-002). https://connect.fsc.org/document-centre/documents/resource/207. Last visited 2022-03-22.

FSC (2018a). FSC Global Market Survey Report 2018. https://fsc.org/en/media/5314. Last visited 2021-09-23.

FSC (2018b). The Share of Sustainable Wood: Data on FSC's Presence in Global Wood Production. https://web.archive.org/web/20220224113603/https://fsc.org/en/newsfeed/the-share-of-sustainable-wood-data-on-fscs-presence-in-global-wood-production. Last visited 2022-02-24.

FSC (2019). FSC Evaluates the Credibility and Integrity of its Assurance System. https://fsc.org/en/newsfeed/fsc-evaluates-the-credibility-and-integrity-of-its-assurance-system. Last visited 2021-09-23.

FSC (2020). Public Certificate Search. https://info.fsc.org/certificate.php#result. Last visited 2021-09-23.

FSC (2021a). FSC statement on Earthsight report 2021. https://fsc.org/en/newscentre/integrity-and-disputes/fsc-statement-on-earthsight-report-2021. Last visited 2023-08-11.

FSC (2021b). Interpretations of the normative framework - Forest Management. https://fsc.org/en/document-centre. Last visited 2021-05-31.

FSC (2021c). Interpretations of the normative framework - General. https://fsc.org/en/document-centre. Last visited 2022-07-26.

FSC (2023a). Document Centre. https://fsc.org/en/document-centre. Last visited 2022-07-26.

FSC (2023b). Public Certificate Search. https://info.fsc.org/certificate.php#result. Last visited 2023-06-20.

FSC and IPSOS (2023). FSC Consumer Recognition Study - Global Report June 2023. https://fsc.org/sites/default/files/2023-06/FSC%20Consumer%20Recognition%20Study%20-%20Global%20Report%20June%202023.pdf. Last visited 2023-09-27.

FSC Watch (2020). FSC Watch. https://fsc-watch.com/. Last visited 2020-05-31.

Galati, A., Gianguzzi, G., Tinervia, S., Crescimanno, M., and Veca, D. S. L. M. (2017). Motivations, adoption and impact of voluntary environmental certification in the Italian Forest based industry: The case of the FSC standard. *Forest Policy and Economics*, 83:169–176.

Gandhi, A. and Houde, J.-F. (2019). Measuring substitution patterns in differentiated products industries. NBER Working Paper 26375, National Bureau of Economic Research.

Gavrilut, I., Halalisan, A.-F., Giurca, A., and Sotirov, M. (2015). The Interaction between FSC Certification and the Implementation of the EU Timber Regulation in Romania. *Forests*, 7(12):3.

Goodman, R. C., Harman Aramburu, M., Gopalakrishna, T., Putz, F. E., Gutiérrez, N., Mena Alvarez, J. L., Aguilar-Amuchastegui, N., and Ellis, P. W. (2019). Carbon emissions and potential emissions reductions from low-intensity selective logging in southwestern Amazonia. *Forest Ecology and Management*, 439:18–27.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo Maximum Likelihood Methods: Applications to Poisson Models. *Econometrica*, 52(3):701.

Greenpeace International (2021). Destruction: Certified. Technical report. Last visited 2021-11-18.

Gutierrez Garzon, A. R., Bettinger, P., Siry, J., Abrams, J., Cieszewski, C., Boston, K., Mei, B., Zengin, H., and Yeşil, A. (2020). A Comparative Analysis of Five Forest Certification Programs. *Forests*, 11(8):863.

Hansen, B. (2021). Econometrics. https://www.ssc.wisc.edu/bhansen/econometrics/. Last visited 2021-09-23.

Hastie, T., Friedman, J., and Tibshirani, R. (2001). *The Elements of Statistical Learning*. Springer Series in Statistics. Springer, New York, NY.

Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica : Journal of the Econometric Society*, 47(1):153–161.

Holopainen, J., Toppinen, A., and Perttula, S. (2015). Impact of European Union Timber Regulation on Forest Certification Strategies in the Finnish Wood Industry Value Chain. *Forests*, 6(12):2879–2896.

Hubbard, T. N. (2002). How do consumers motivate experts? Reputational incentives in an auto repair market. *The Journal of Law and Economics*, 45(2):437–468.

Hui, X., Saeedi, M., Spagnolo, G., and Tadelis, S. (2023). Raising the Bar: Certification Thresholds and Market Outcomes. *American Economic Journal: Microeconomics*, 15(2):599–626.

IMF (2020). World Economic Outlook Database. https://www.imf.org/en/Publications/WEO/weo-database/2020/April. Last visited 2023-06-20.

ISO (2023). ISO 9000 family - Quality management. https://www.iso.org/iso-9001-quality-management.html. Last visited 2023-10-03.

Jaung, W., Putzel, L., Bull, G. Q., Kozak, R., and Elliott, C. (2016). Forest Stewardship Council certification for forest ecosystem services: An analysis of stakeholder adaptability. *Forest Policy and Economics*, 70:91–98.

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

Kalonga, S. K., Midtgaard, F., and Klanderud, K. (2016). Forest certification as a policy option in conserving biodiversity: An empirical study of forest management in Tanzania. *Forest Ecology and Management*, 361:1–12.

Kitchoukov, E., Stoyanov, N., Simeonova-Zarcin, T., and Stoyanov, T. (2019). Profitability of Forest Certification – Case Study Analysis. *Economic Alternatives*, (4):607–614.

Klein, B. and Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89(4):615–641.

Kollert, W. and Lagan, P. (2007). Do certified tropical logs fetch a market premium? *Forest Policy and Economics*, 9(7):862–868.

Landsberger, H. A. (1958). *Hawthorne Revisited: Management and the Worker, Its Critics, and Developments in Human Relations in Industry.* ERIC.

Lee, L.-F. (1983). Generalized econometric models with selectivity. *Econometrica : Journal of the Econometric Society*, 51(2):507–512.

Ludwig, G., Tronicke, C., Köck, W., and Gawel, E. (2014). Rechtsrahmen der Bioökonomie in Mitteldeutschland - Bestandsaufnahme und Bewertung. UFZ Discussion Papers 22, Leipzig: Helmholtz-Zentrum für Umweltforschung - UFZ.

Mathis, J., McAndrews, J., and Rochet, J.-C. (2009). Rating the raters: Are reputation concerns powerful enough to discipline rating agencies? *Journal of Monetary Economics*, 56(5):657–674.

Mayer, T. and Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database. CEPII Working Paper 2011-25.

McFadden, D. (1977). Modeling the choice of residential location. Cowles Foundation Discussion Papers 710.

Miteva, D. A., Loucks, C. J., and Pattanayak, S. K. (2015). Social and Environmental Impacts of Forest Management Certification in Indonesia. *PLoS ONE*, 10(7):e0129675.

NEPCon (2019). FSC TM Forest Management certification - Service information. www.nepcon.org. Last visited 2019-12-02.

Nordhaus, W. D. (2017). Revisiting the social cost of carbon. *Proceedings of the National Academy of Sciences*, 114(7):1518–1523.

Overdevest, C. and Rickenbach, M. G. (2006). Forest certification and institutional governance: An empirical study of forest stewardship council certificate holders in the United States. *Forest Policy and Economics*, 9(1):93–102.

Owari, T. and Sawanobori, Y. (2007). Analysis of the certified forest products market in Japan. *Holz als Roh- und Werkstoff*, 65(2):113–120.

Paluš, H., Krahulcová, M., and Parobek, J. (2021). Assessment of Forest Certification as a Tool to Support Forest Ecosystem Services. *Forests*, 12(3):300.

Panlasigui, S., Rico-Straffon, J., Pfaff, A., Swenson, J., and Loucks, C. (2018). Impacts of certification, uncertified concessions, and protected areas on forest loss in Cameroon, 2000 to 2013. *Biological Conservation*, 227:160–166.

Preston, A. E. (1989). The Nonprofit Worker in a For-Profit World. *Journal of Labor Economics*, 7(4):438–463.

Rafael, G. C., Fonseca, A., and Jacovine, L. A. G. (2018). Non-conformities to the Forest Stewardship Council (FSC) standards: Empirical evidence and implications for policy-making in Brazil. *Forest Policy and Economics*, 88:59–69.

Rico-Straffon, J., Wang, Z., Panlasigui, S., Loucks, C. J., Swenson, J., and Pfaff, A. (2023). Forest concessions and eco-certifications in the Peruvian Amazon: Deforestation impacts of logging rights and logging restrictions. *Journal of Environmental Economics and Management*, 118:102780.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Shim, E. and Sudit, E. F. (1995). How manufacturers price products. *Strategic Finance*, 76(8).

Simonet, G., Subervie, J., Ezzine-de Blas, D., Cromberg, M., and Duchelle, A. E. (2019). Effectiveness of a REDD+ Project in Reducing Deforestation in the Brazilian Amazon. *American Journal of Agricultural Economics*, 101(1):211–229.

Sims, K. R. and Alix-Garcia, J. M. (2017). Parks versus PES: Evaluating direct and incentive-based land conservation in Mexico. *Journal of Environmental Economics and Management*, 86:8–28.

Souza-Rodrigues, E. (2019). Deforestation in the Amazon: A Unified Framework for Estimation and Policy Analysis. *The Review of Economic Studies*, 86(6):2713–2744.

Stahl, K. and Strausz, R. (2017). Certification and Market Transparency. *Review of Economic Studies*, 84(4):1842–1868. https://search-ebscohost-com-s.biblio-dist.ut-capitole.fr/login.aspx?direct=true&db=eoh&AN=1686895&lang=fr&site=ehost-live.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. eds Donald W.K. Andrews and James H. Stock, Cambridge University Press.

Subervie, J. and Vagneron, I. (2013). A Drop of Water in the Indian Ocean? The Impact of GlobalGap Certification on Lychee Farmers in Madagascar. *World Development*, 50:57–73.

Sugiura, K. and Oki, Y. (2018). Reasons for Choosing Forest Stewardship Council (FSC) and Sustainable Green Ecosystem Council (SGEC) Schemes and the Effects of Certification Acquisition by Forestry Enterprises in Japan. *Forests*, 9(4):173.

Terza, J. V. (1998). Estimating count data models with endogenous switching: Sample selection and endogenous treatment effects. *Journal of Econometrics*, 84(1):129–154.

Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Transparency International (2021). Corruption Perception Index. https://images.transparencycdn.org/images/CPI_FULL_DATA_2021-01-27-162209.zip. Last visited 2021-09-23.

Tritsch, I., Le Velly, G., Mertens, B., Meyfroidt, P., Sannier, C., Makak, J.-S., and Houngbedji, K. (2020). Do forest-management plans and FSC certification help avoid deforestation in the Congo Basin? *Ecological Economics*, 175(106660).

UNECE and FAO (2023). Prices. https://unece.org/forests/prices. Last visited 2023-08-13.

UNECE/FAO (2019). Forest Products Annual Market Review, 2018-2019. Technical report, United Nations Economic Commission for Europe (UNECE) and the Food and Agriculture Organization of the United Nations (FAO).

United Nations (2021). Global SDG Indicators Database. https://unstats.un.org/sdgs/indicators/database/. Last visited 2021-09-23.

United Nations (2023). UNSD — Methodology. https://unstats.un.org/unsd/methodology/m49/. Last visited 2023-10-03.

Villalobos, L., Coria, J., and Nordén, A. (2018). Has forest certification reduced forest degradation in Sweden? *Land Economics*, 94(2):220–238.

Visseren-Hamakers, I. J. and Pattberg, P. (2013). We Can't See the Forest for the TreesThe Environmental Impact of Global Forest Certification Is Unknown. *GAIA-Ecological Perspectives for Science and Society*, 22(1):25–28.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, Mass, 2nd ed edition.

World Bank (2020). DEC alternative conversion factor (LCU per US$). https://data.worldbank.org/indicator/PA.NUS.ATLS. Last visited 2021-09-23.

World Bank (2021). World Bank Open Data. https://data.worldbank.org/. Last visited 2021-09-01.

Zeileis, A., Fox, J., and Kleber, A. (2023). Ivreg: Instrumental-Variables Regression by '2SLS', '2SM', or '2SMM', with Diagnostics. https://zeileis.github.io/ivreg/. Last visited 2023-09-26.