RIEEM Research Institute for Environmental Economics and Management

Discussion Paper Series No.2503

Do Eco-labels Pay Off? Causal Evidence from Japanese Firms

Shigeharu Okajima , Hiroko Okajima , Naohiro Shirao & Kenji Takeuch

January 2026

WASEDA

Do Eco-labels Pay Off? Causal Evidence from Japanese Firms

Shigeharu Okajima[1], Hiroko Okajima[2], Naohiro Shirao[3], and Kenji Takeuchi[4]

Abstract

Eco-labels are widely promoted as information-based environmental instruments that generate "win–win" outcomes by improving both environmental quality and firm profitability. However, credible causal evidence on their financial effects remains limited. Using panel data on Japanese firms from 2012 to 2016, this study examines whether Type I (third-party certified) and Type II (self-declared) eco-labels improve firm financial performance. To address selection bias, we apply inverse probability weighting with firm and year fixed effects and stabilize weights through trimming and capping procedures. We further examine heterogeneity between B2C and B2B firms based on differences in consumer visibility. The results show that Type I labels have no significant financial effects across all specifications. Type II labels exhibit modest positive effects for B2C firms under trimmed weights, but these effects disappear when extreme weights are capped, indicating limited robustness. Overall, we find no consistent financial benefits from eco-label adoption, challenging the business-case narrative and suggesting that eco-label policies should be justified primarily by environmental effectiveness rather than expected profitability gains.

Keywords:

Eco-labels; Environmental labeling; Firm performance; Causal inference; Inverse probability weighting

JEL Classification Codes:

Q50; Q58; M14

---

[1] Kobe University, Graduate School of International Cooperation Studies, 2-1 Rokkodai-cho, Nada-ku, Kobe 657-8501, Email: Shigeharu.okajima@gmail.com

[2] Nagoya University, Nagoya University Graduate School of Economics, Furocho, Chikusa Ward, Nagoya City Aichi 464-8601, Email: hiroko.okajima@gmail.com

[3] Osaka University of Economics, 2-2-8 Osumi HigashiYodogawa-ku Osaka-shi, 533-8533
Email: bozu0610bozu@gmail.com

[4] Kyoto University, Graduate School of Global Environmental Studies, Yoshida-honmachi Kyoto,606-8501, Email: takeuchi@econ.kyoto-u.ac.jp

1. Introduction

Eco-labels are prominent tools in information-based environmental policy. Global directories list over 450 eco-label schemes in 199 countries across 25 industry sectors, reflecting the rapid expansion of programs that communicate product-level environmental attributes (Ecolabel Index; Meis-Harris et al. 2021). By providing standardized environmental information on product packaging, eco-labels aim to reduce information asymmetry, shift consumer demand toward greener products, and incentivize firms to adopt environmentally responsible practices.

Among environmental certifications—such as environmental management systems, carbon disclosure initiatives, and ESG ratings—eco-labels hold a uniquely visible and intuitive position. They translate complex environmental attributes into recognizable signals that influence consumer decisions at the point of purchase. Due to this consumer-facing visibility, governments and firms have heavily invested in eco-label infrastructure, including certification standards, auditing procedures, and public information programs (OECD 2016).

A substantial empirical literature examines whether eco-labels deliver economic value. Early studies report that eco-labels generate price premiums and increase consumer willingness to pay (Nimon and Beghin 1999; Cason and Gangadharan 2002; Teisl et al. 2008). Subsequent work indicates that eco-labels influence demand patterns and market shares (Brécard et al. 2009; Schleenbecker and Hamm 2013). Recent firm-level evidence suggests that eco-labeled products may improve financial performance (Schweizer and Zellweger 2022).

However, the policy relevance of this literature is limited for two reasons. First, firms adopting eco-labels differ systematically from those that do not. Eco-label adopters tend to be larger, more profitable, better managed, and more committed to CSR—all characteristics that independently influence financial performance. This non-random selection complicates attributing performance differences to eco-label adoption rather than pre-existing firm attributes. This issue is recognized in the broader literature on voluntary environmental programs, where substantial self-selection into certification can bias estimates of program effects (King and Lenox 2001; Darnall and Sides 2008).

Most empirical studies on eco-labels do not adequately address selection bias. Demand-side research—examining consumer willingness to pay or choice behavior—typically relies on OLS, probit models, or controlled experiments where selection concerns are minimized by design. Conversely, firm-level studies assessing the financial impacts of eco-labels are scarce and primarily correlational. For instance, Schweizer and Zellweger (2022) document positive associations between labeling and performance but do not isolate exogenous variation in certification. Environmental economics research emphasizes the importance of credible causal inference and rigorous empirical designs when evaluating environmental practices (Greenstone et al., 2012). In this study, we contribute by moving beyond simple correlations and applying a state-of-the-art observational causal framework that addresses selection on observables using inverse probability weighting and fixed effects, while acknowledging that quasi-experimental designs could further strengthen identification.

Additionally, the literature often treats eco-labels as a homogeneous category despite substantial institutional heterogeneity. Under the ISO framework, environmental labels differ markedly in their verification requirements and informational content, most notably between third-party certified labels (Type I) and self-declared environmental claims (Type II). Signaling theory suggests that only credible and externally verified signals can effectively reduce information asymmetry (Spence, 1973), yet few empirical studies compare these label types within a unified causal framework. This study addresses this gap by explicitly distinguishing between Type I and Type II eco-labels.

This study provides causal evidence on the financial effects of eco-labels by analyzing panel data from Japanese firms between 2012 and 2016. Our contribution is twofold. First, we estimate treatment effects using inverse probability weighting combined with firm and year fixed effects (IPW–FE). To ensure covariate balance, we trim extreme weights at the 99th percentile and apply capping procedures as robustness checks. Second, we compare Type I and Type II labels within a unified empirical framework to test whether label credibility—operationalized through third-party verification—affects financial returns.

We also examine heterogeneous treatment effects between B2C and B2B firms. Eco-labels primarily aim to influence consumer perceptions, suggesting a stronger financial impact for firms selling directly to consumers than for upstream suppliers. Our analysis assesses whether market position and consumer visibility affect the financial implications of eco-label adoption.

Japan is an ideal setting for evaluating these questions for three reasons. First, both Type I (Eco Mark) and Type II self-declared labels coexist within a mature institutional framework, allowing for a direct comparison of credible versus non-credible environmental signals. Second, the CSR Database offers rich and consistent panel data on firm-level label adoption, enabling us to trace within-firm changes over time. Third, consumer awareness of environmental labeling in Japan is relatively high, ensuring that any null effects cannot be attributed to a lack of visibility. Section 2 elaborates on these institutional features.

Our findings reveal that neither Type I nor Type II eco-labels generate statistically significant improvements in financial performance. Even third-party certified labels, presumed to provide credible signals, do not yield measurable financial gains. These results challenge the assumption that eco-labels create “win–win” outcomes, benefiting both the environment and firm profitability. For policymakers, these findings emphasize the need to justify eco-label programs based on environmental effectiveness rather than presumed financial incentives.

The remainder of this study is structured as follows. Section 2 describes the institutional background of eco-labels in Japan. Section 3 outlines the empirical strategy, including the IPW–FE framework and weight-stabilization procedures. Section 4 reports the main results, including heterogeneity analyses. Section 5 discusses mechanisms and implications. Section 6 concludes.

2. Institutional Background

2.1 Environmental Labeling in Japan

Japan has been a global pioneer in environmental labeling since the late 1980s, demonstrating a commitment to environmental governance through regulatory initiatives and voluntary industry programs. Japanese consumers exhibit high environmental awareness, and firms are incentivized to signal their environmental responsibility through visible product certifications. Japan's eco-label system aligns with the ISO framework and includes two major categories:

- Type I (third-party certified) labels requiring independent verification based on life-cycle criteria;
- Type II (self-declared) labels that firms may adopt voluntarily without external auditing.

Both label types are prevalent in manufacturing, retail, and services, creating a unique market where credible and less credible environmental signals coexist.

2.2 Type I Eco-Labels: The Eco Mark Program

The Eco Mark, launched in 1989 by the Japan Environment Association (JEA), is Japan's primary Type I eco-label and one of the world's earliest national labeling programs. Key features include:

- Life-cycle–based criteria covering raw material extraction, manufacturing, distribution, use, and disposal;
- Third-party auditing prior to certification;
- Periodic renewal to ensure continued compliance;
- Category-specific standards, currently covering over 50 product groups.

As of 2014, approximately 5,553 products across 59 categories were Eco Mark certified (UN Environment Programme, 2018).

Despite its credibility, the program faces challenges: certification and compliance costs are substantial, particularly for small and medium-sized enterprises, and increased adoption may dilute the signaling value of certification as differentiation becomes harder.

2.3 Type II Eco-Labels: Self-Declared Environmental Claims

Type II labels comprise voluntary firm-initiated claims such as "eco-friendly," "recyclable," or "energy-saving." These claims are governed by ISO 14021 guidelines, which advise that statements be accurate, verifiable, and non-misleading. However, enforcement is limited, and firms face few penalties for vague or unsubstantiated claims. The low cost of adoption makes Type II labels appealing to many firms; however, the absence of independent verification raises concerns about credibility. Surveys indicate growing consumer skepticism toward self-declared green claims, particularly amid increased awareness of greenwashing (Consumers International, 2022).

2.4 Regulatory Environment and Disclosure

Japan's institutional environment provides rich, structured data on environmental practices. The CSR Database, compiled annually by Toyo Keizai, collects detailed information on environmental management, including eco-label adoption, for listed and major unlisted firms. This ensures consistent panel data coverage that is rare in other countries. Additionally, the Green Purchasing Law (2000) encourages public agencies to prioritize environmentally preferable products, increasing demand for eco-labeled goods in government procurement and influencing firms' incentives to obtain credible Type I certification.

2.5 Why Japan Provides an Advantageous Empirical Setting

Japan offers an ideal setting for this analysis due to several factors. The coexistence of certified and self-declared labels enables a clear comparison of signal credibility within a single institutional environment. Adoption rates are substantial: approximately 14% of firms adopt Type I labels, and 11% adopt Type II labels, providing sufficient cross-sectional and within-firm variation to assess whether the informational value of eco-labels diminishes with widespread labeling. Rich panel data from the CSR Database supply the within-firm variation necessary for credible causal inference. The country's eco-label institutions have been established for over three decades, allowing for the evaluation of long-run equilibrium effects. Furthermore, consumer awareness of environmental labeling is high, ensuring that any null effects cannot be attributed to limited visibility or recognition.

2.6 Theoretical Framework and Research Questions

This study draws on signaling theory (Spence 1973) to examine whether eco-labels generate financial returns for firms. Signaling theory posits that credible and costly-to-fake signals effectively reduce information asymmetry and influence firm outcomes. In the context of environmental labeling, varying verification requirements and credibility suggest divergent financial effects across label types.

Type I vs. Type II Labels

Type I eco-labels require third-party certification based on standardized, life-cycle-based criteria, making them relatively costly and difficult to imitate. These characteristics enhance credibility and should, in principle, strengthen the label's signaling value. In contrast, Type II eco-labels consist of self-declared environmental claims without independent verification. Their low cost and ease of adoption raise concerns about credibility, "greenwashing," and weaker informational content. If signaling theory holds, Type I labels should yield stronger financial effects than Type II labels.

However, signaling value is context-dependent. In markets with high adoption rates, even credible signals may lose their differentiating power. Japan's widespread adoption of eco-labels (Sections 2.2–2.3) raises the possibility of saturation effects that may diminish both Type I and Type

II signals, regardless of their inherent credibility.

Consumer Visibility and Market Position

Eco-labels are designed to influence consumer perceptions, implying their financial impact depends on a firm's position in the value chain. B2C firms are more likely to benefit from environmental signaling, as labels are visible at the point of purchase. In contrast, upstream or B2B firms have limited consumer visibility, suggesting eco-labels may exert weaker or negligible financial effects. This motivates an examination of heterogeneous treatment effects by firm type.

Research Questions

Drawing on these theoretical considerations, the study investigates the following research questions:

1. Do Type I (third-party certified) eco-labels improve firm financial performance?
2. Do Type II (self-declared) eco-labels improve firm financial performance?
3. Do the effects of eco-label adoption differ between B2C and B2B firms?

Given the conflicting predictions from signaling theory (which suggests Type I labels should generate stronger financial effects) and market saturation theory (which suggests both types may fail in mature labeling environments), and the limited availability of causal evidence in prior research, we frame these questions as open empirical inquiries rather than pre-specifying directional hypotheses. This approach is methodologically appropriate where theoretical mechanisms diverge, and empirical knowledge is limited. It allows the data to reveal patterns that rigid ex-ante hypotheses might obscure while maintaining transparency regarding theoretical ambiguity.

3. Empirical strategy

3.1 Data

Firm-level eco-label adoption data are obtained from the CSR Database maintained by Toyo Keizai for the period 2012–2016. This annual survey covers listed and major unlisted firms and reports whether firms use Type I labels, Type II labels, both, or neither. Although eco-labels are assigned at the product level, the CSR survey records firm-level adoption across product portfolios, which constitutes the most consistent and widely used measure in empirical research on eco-labels (Schweizer and Zellweger, 2022; Arimura et al., 2011).

To separately identify the effects of different labeling schemes, we construct mutually exclusive treatment groups and focus on firms that adopt only one label type. Specifically, firms are classified as (i) Type I only adopters, (ii) Type II only adopters, or (iii) non-adopters. Firms adopting both label types are excluded from the analysis. Financial outcomes—including return on assets (ROA), operating profit, and Tobin's Q—as well as industry codes, leverage, employment, and other control variables are obtained from the Nikkei NEEDS Financial QUEST database. The resulting panel provides substantial within-firm variation, although label adoption is relatively persistent over time, underscoring the importance of causal identification strategies that account for pre-existing firm heterogeneity.

Although the ISO classification distinguishes three types of environmental labels, our empirical analysis focuses on Type I and Type II labels. Type III environmental product declarations (EPDs) account for only 0.9% of firm-year observations in our dataset and are primarily used as B2B disclosure tools rather than consumer-facing signals. This extremely limited adoption precludes meaningful statistical analysis and prevents reliable causal inference. We therefore exclude Type III labels from the analysis and concentrate on the two label types with sufficient variation and relevance for firm-level performance evaluation.

Choice of Outcome Variables

This study examines four outcome variables that capture complementary dimensions of firm performance. These measures allow us to distinguish between market-based valuation effects and accounting-based operational outcomes.

(1) Tobin's Q.

Tobin's Q reflects market valuation and investor expectations regarding future profitability and intangible asset value.

(2) Operating profit (log).

This measure captures core operational profitability and short-run earnings performance.

(3) Operating ratio.

Defined as:

$$\text{Operating Ratio}_{it} = \frac{\text{Operating Profit}_{it}}{\text{Sales}_{it}} \times 100,$$

the operating ratio provides an accounting-based indicator of operational efficiency that complements the log operating profit measure.

(4) ROA.

ROA reflects the efficiency with which firms convert assets into net operating returns:

$$\text{ROA}_{it} = \frac{\text{Operating Profit}_{it}}{\text{Total Assets}_{it}} \times 100,$$

where total assets are measured in hundreds of millions of yen.

Together, these indicators offer a multidimensional assessment of firm performance, enabling us to evaluate whether eco-label adoption affects market valuation, operational profitability, or asset efficiency.

Covariate construction.

The empirical analysis relies on a set of firm-level covariates commonly used in studies of voluntary environmental programs and corporate environmental behavior. These variables capture financial capacity, operational scale, and market power—factors that influence both eco-label adoption and firm performance.

(1) Financial capacity and leverage.

We include the debt ratio, defined as interest-bearing liabilities divided by total assets:

$$\text{Debt Ratio}_{it} = \frac{\text{Interest-Bearing Debt}_{it}}{\text{Total Assets}_{it}}.$$

This measure reflects financial constraints and borrowing capacity, which may influence a firm's ability to undertake certification-related investments.

(2) Firm size.

Firm scale is measured by the number of employees (in thousands). Larger firms may be more capable of absorbing certification costs or engaging in environmental disclosure activities.

(3) Market power.

Market power is proxied by a profit-margin measure constructed following the spirit of Aghion et al. (2005):

$$\text{Profit margin}_{it} = \frac{\text{Operating Profit}_{it} - \text{Interest Payments}_{it}}{\text{Sales}_{it}}.$$

While this measure reflects pre-tax profitability rather than a pure operating margin, it captures firm-level pricing power and competitive conditions faced by firms. We use this proxy as a control variable rather than as a structural measure of markups.

Descriptive statistics for the three mutually exclusive groups—non-adopters, Type I–only adopters, and Type II–only adopters—are presented in Table 1. The table shows systematic differences between eco-label adopters and non-adopters across key characteristics. Adopting firms tend to be larger, more profitable, and exhibit higher average markups, indicating substantial non-random selection into eco-label adoption. Pairwise mean-comparison tests (Bonferroni-adjusted) confirm these patterns: Type I and Type II adopters have significantly higher employee counts, sales, and operating profits than non-adopters, while differences in Tobin's Q and ROA are statistically insignificant. Detailed results are reported in Appendix Table A1.

We employ inverse-probability weighting (IPW) based on separately estimated propensity score models for Type I and Type II labels to adjust for observable differences between adopters and non-adopters.

Table 1: Descriptive Statistics for Control, Type I Only, and Type II Only Firms

Variable	Non-adopters	Type I only	Type II only
	0.195	0.191	0.183
Debt ratio	(0.159)	(0.153)	(0.149)
	Obs = 2,132	Obs = 388	Obs = 320
	9.074	16.587	22.895
Employees (×1,000)	(26.626)	(44.087)	(40.823)
	Obs = 1,903	Obs = 368	Obs = 305
	0.052	0.107	0.063
Profit margin	(0.181)	(0.190)	(0.175)
	Obs = 1,610	Obs = 278	Obs = 238
	189,698	503,543	293,218
Sales (million yen)	(471,303)	(1,455,530)	(456,109)
	Obs = 1,612	Obs = 282	Obs = 238
	1.159	1.170	1.231
Tobin's Q	(0.918)	(0.547)	(0.571)
	Obs = 2,132	Obs = 388	Obs = 320
	7,984	30,501	20,784
Operating profit (million yen)	(33,572)	(145,062)	(61,912)
	Obs = 1,610	Obs = 278	Obs = 238
	2.911	2.848	2.257
ROA (%)	(4.933)	(3.106)	(3.230)
	Obs = 1,610	Obs = 278	Obs = 238

Notes: Table reports descriptive statistics for the three mutually exclusive groups of firms: non-adopters, Type I adopters, and Type II adopters. Means are shown with standard deviations in parentheses. Sample sizes vary across variables because several financial-statement items are not reported for all firm-year combinations in NEEDS Financial QUEST. Employee counts and Profit margin values are unavailable for some firms, resulting in sample-size differences of approximately 20-30%. This also explains outcome-specific sample sizes in Tables 4-6.

We focus on the 2012–2016 period for three reasons. First, the CSR Database offers complete and consistent coverage during these years, ensuring high-quality panel continuity across firms. Second, this window captures a mature phase of eco-label adoption in Japan, avoiding the early diffusion period when selection dynamics, motives, and regulatory expectations varied significantly. Third, this period is free from major structural breaks or policy shocks—such as the post-2011 recovery phase or the implementation of the Paris Agreement after 2016—that could confound treatment effect estimation. Limiting the analysis to this stable period enhances internal validity, ensuring that estimated treatment effects reflect firm-level responses rather than macro-level disruptions.

3.2 Identification Strategy

The descriptive statistics in Table 1 reveal significant differences between adopters and non-adopters, raising concerns about selection bias. Firms that adopt eco-labels are larger, more profitable, and possess greater market power—characteristics that independently affect financial performance. These systematic differences suggest that simple comparisons between treated and untreated firms do not yield credible estimates of causal effects.

To address this selection problem, we combine inverse probability weighting (IPW) with firm and year fixed effects (FE). IPW reweights the sample to balance observable characteristics between treated and control firms, approximating a pseudo-population where treatment assignment is independent of observed covariates (Busso et al. 2014). Firm fixed effects control for time-invariant unobserved heterogeneity, such as managerial quality or organizational culture, while year fixed effects account for macroeconomic shocks. This IPW–FE strategy effectively addresses both observable and unobservable selection components that remain stable over time and has demonstrated efficacy in observational settings with substantial selection (Abadie and Imbens 2016).

3.2.1 Propensity Score Estimation

To address selection bias in eco-label adoption, we estimate propensity scores (PS) separately for Type I (third-party certified) and Type II (self-declared) eco-labels.

For each labeling scheme, we estimate a logit model of the form:

$$\hat{p}(X_{it}) = \Pr\,(D_{it} = 1 \mid X_{it}),$$

where D_{it} is the adoption indicator and X_{it} includes key observable characteristics associated with firms' environmental management capacity and financial performance. Here, i denotes firms and t denotes fiscal years. The covariate vector X_{it} comprises three key firm characteristics: the debt ratio, the number of employees (in thousands), and the Profit margin. These variables are standard predictors in the literature on voluntary environmental programs and corporate environmental behavior.

The propensity score model is deliberately kept parsimonious. In the IPW–FE framework, selection on time-invariant firm characteristics is addressed by firm fixed effects, while IPW adjusts for selection on observable, time-varying factors. We therefore include only key time-varying predictors of eco-label adoption that are standard in the literature and plausibly predetermined with respect to eco-label adoption (e.g., Busso et al., 2014; Abadie and Imbens, 2016).

This study does not compare Type I and Type II adopters directly. The two schemes differ fundamentally in institutional design—Type I labels involve third-party certification, while Type II labels are self-declared environmental claims. Firms adopt these labels for distinct reasons; thus, Type I and Type II adopters do not belong to a unified choice set suitable for a multinomial treatment framework. We estimate separate binary propensity-score models for each adoption margin (Type I only vs. non-adopters; Type II only vs. non-adopters), which allows us to adjust for observable selection without imposing the restrictive assumption that firms choose between both label types simultaneously.

A key feature of the data is the highly uneven distribution of estimated propensity scores, leading to heavy-tailed IPW weights. This occurs for three institutional reasons. First, both Type I and Type II adopters draw from the same non-adopter pool, while adoption rates for each label type are low, resulting in very small predicted probabilities for many treated firms. Second, some firms have covariate profiles that differ significantly from the control group, creating near-deterministic treatment probabilities for a few observations. Third, since the binary PS models are estimated separately for each label type, the effective support for each adoption margin is narrow. Together, these factors contribute to extreme weights, necessitating trimming and capping in subsequent analyses.

Using the predicted propensity scores, we construct inverse-probability weights (IPW) for each observation:

$$w_{it} = \frac{D_{it}}{\hat{p}(X_{it})} + \frac{1 - D_{it}}{1 - \hat{p}(X_{it})}.$$

These weights reweight the sample such that the distribution of covariates in the treated and control groups becomes comparable.

3.3 Diagnosing Common Support and Weight Instability

A critical requirement for IPW is the overlap (or practical positivity) condition: treated and control firms must exhibit similar ranges of estimated propensity scores. Figure 1 presents kernel density plots of the propensity score distributions for treated and control firms in both the Type I and Type II samples. While a significant share of observations lies within a common region, both cases show meaningful subsets of control firms with estimated treatment probabilities very close to one. This evidence indicates limited common support and raises concerns about the empirical stability of the IPW estimator.

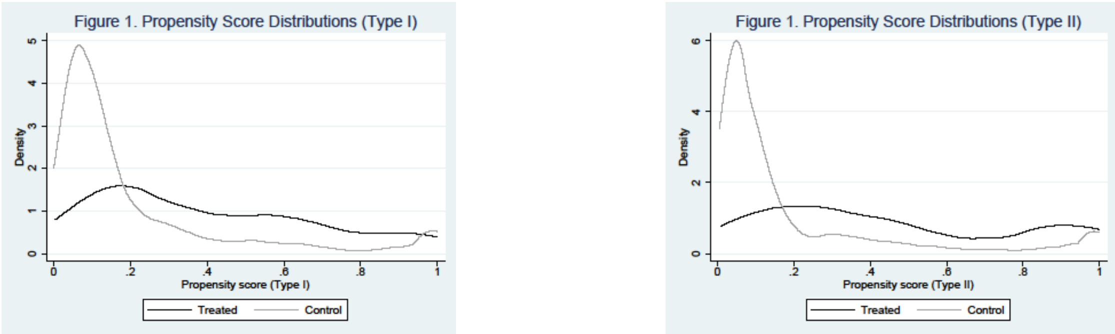

Figure 1: Propensity Score Distributions for Type I and Type II Eco-Labels

These concerns are supported by the raw IPW weight distributions in Table 2. The distributions show extreme heavy-tailedness: for Type I adoption, the 99th percentile exceeds 33,000, and the maximum surpasses 16 million; for Type II adoption, these values exceed 170,000 and 16.8 million, respectively. Such values indicate that a small number of observations receive disproportionately large weights, violating the practical positivity assumption and causing the estimator to rely heavily on a few firms (Busso et al. 2014; Abadie and Imbens 2016).

Table 2: Raw IPW Weight Distribution for Type I and Type II Adoption

Percentile	Type I Adoption	Type II Adoption
90th percentile	4.47	4.71
95th percentile	13.12	13.25
99th percentile	33,893	174,763
Maximum	> 16,000,000	16,800,000

Notes: Percentiles are calculated from raw inverse-probability weights. Extreme maximum values reflect limited practical positivity and the presence of a small number of observations with estimated treatment probabilities approaching one.

Despite extreme weight instability, the propensity score models substantially improve covariate balance. Before weighting, raw standardized differences for debt ratio, employment size, and the Profit margin range from 0.02 to 0.29, indicating significant pre-treatment imbalance. After applying IPW, standardized differences are markedly reduced, and variance ratios fall within accepted ranges (0.8–1.25 for Type I and 0.917–1.045 for Type II), indicating satisfactory covariate balance in both samples (Tables 3A and 3B). While some covariates in the Type II sample retain moderate standardized differences, these are substantially smaller than in the raw data.

However, effective covariate balance does not address the severe heavy-tailedness of raw weights. As Type I and Type II treatment groups do not overlap—404 firms adopt only Type I labels and 323 adopt only Type II labels—both specifications draw from the same control pool (N = 2,177), resulting in similar balance patterns but distinct weight distributions. The extreme concentration of raw weights necessitates stabilization through trimming or capping to ensure reliable inference in subsequent analyses.

Table 3A: Covariate Balance for Type I Propensity Score Model

Covariate	Raw Std. Diff	Weighted Std. Diff	Raw Var. Ratio	Weighted Var. Ratio
debt	–0.023	–0.000	0.98	1.028
employees (×1000)	0.136	–0.016	2.015	1.045
Profit margin	0.293	–0.014	1.362	0.917

Table 3B: Covariate Balance for Type II Propensity Score Model

Covariate	Raw Std. Diff	Weighted Std. Diff	Raw Var. Ratio	Weighted Var. Ratio
debt	0.074	0.011	0.916	0.979
employees (×1000)	0.379	–0.136	2.756	0.379
Profit margin	-0.289	0.123	0.861	0.513

3.4 Weight Stabilization: Trimming and Capping

Section 3.3 revealed that raw IPW weights exhibit extreme heavy-tailedness, with maximum values exceeding 16 million. Such weights violate practical positivity, inflate estimator variance, and render the IPW estimator unreliable. To address this issue, we apply two weight-stabilization procedures—trimming and capping—to reduce the influence of large weights while preserving core identifying variation.

Trimming at the 99th percentile

Following Busso et al. (2014) and Crump et al. (2009), we trim observations with propensity-score-based weights above the 99th percentile. These studies demonstrate that extreme weights—resulting from limited overlap—inflate estimator variance and cause the IPW estimator to depend excessively on a small number of observations. Trimming the top 1% is a standard, theoretically justified approach to restore practical positivity. We replace all weights exceeding the 99th percentile threshold with that cutoff value.

$$w_{it}^{\text{trim99}} = \begin{cases} w_{it}, & \text{if } w_{it} \le p_{99}, \\ p_{99}, & \text{if } w_{it} > p_{99}, \end{cases}$$

where p_{99} is computed separately for Type I and Type II models.

This approach retains nearly all observations while removing extreme outliers that violate positivity.

Capping at 10

We cap the weights at 10, following Cole and Hernán (2008) and Austin and Stuart (2015), who recommend bounding extreme IPW weights to reduce variance and prevent undue influence from observations with near-deterministic treatment probabilities. Weight capping at 10 is a widely used robustness procedure in causal inference. Accordingly, we set the maximum weight to 10 in all capped-weight specifications.

$$w_{it}^{\text{cap10}} = \min\left(w_{it}, 10\right).$$

This conservative cap is used in empirical applications with severe overlap problems, significantly reducing estimator variance. After capping, the weight distribution improves: extreme percentiles collapse to the thresholds, variance decreases sharply, and no observation dominates the weighted regression. Re-estimating all outcome models with both sets of stabilized weights yields consistent results, confirming that extreme-weight observations do not drive core findings.

These stabilization procedures ensure the IPW estimator remains reliable under limited common support while preserving the covariate balance achieved by the original PS model.

4. Results

Throughout the analysis, sample sizes vary for three reasons. First, Table 1 reports the full set of firm-year observations, but many accounting variables (operating profit, sales, total assets) are missing for a significant share of firms, leading to different Ns across variables. Second, in the main IPW specifications (Tables 4A and 4B), the effective sample size differs by outcome because each financial indicator requires specific accounting items. Third, the heterogeneity analyses in Tables 5 and 6 further segment the sample into B2C and B2B firms, reducing the number of observations in each subgroup. These factors explain the progression of Ns from Table 1 to Tables 4–6, ensuring that each estimate utilizes all available data for the relevant outcome and subgroup.

4.1 Preliminary Diagnostics

Before addressing the research questions, we assess the suitability of the identification strategy and weighting procedure for causal inference. As shown in Section 3, the propensity score models achieve excellent covariate balance after applying inverse probability weighting (Tables 3A–3B). The weight-stabilization procedures—99th-percentile trimming and capping at 10—eliminate extreme tail behavior in the raw IPW weights, yielding stable weights for both Type I and Type II adoption.

For Type I, trimming collapses the extreme right tail to the 99th-percentile cutoff, while capping produces a tightly bounded distribution with low skewness and kurtosis. Type II weights exhibit similar improvements: trimming removes the pathological upper tail, and capping yields a compact distribution where no single observation has undue influence. These diagnostics confirm that the stabilized IPW estimators provide a solid basis for estimating treatment effects and addressing RQ1–RQ3.

4.2 Stabilizing the IPW Weights: Distributional Effects

4.2.1 Type I Adoption Weights

Trimming at the 99th percentile significantly reduces the impact of extreme observations. Under the trimming rule, only observations above the 99th percentile are affected, while lower percentiles remain unchanged. As a result, the mean weight decreases to 12,242 and the standard deviation to 16,261. The 90th and 95th percentiles remain close to their raw values, whereas the maximum weight equals the trimming threshold of 33,893, indicating that the pathological right tail has been effectively removed.

Capping the weights at 10 yields an even more conservative distribution. The mean falls to 4.89, the standard deviation to 4.23, and all upper percentiles compress to 10. Skewness (0.33) and kurtosis (1.16) are extremely low, reflecting a tightly bounded distribution that prevents any observation from exerting undue influence.

4.2.2 Type II Adoption Weights

The Type II weights show similar improvements after stabilization. Trimmed weights have a mean of 69,025 and a standard deviation of 85,398, with the 90th–99th percentiles all equal to 174,763, the 99th percentile cutoff. As with Type I, the extreme right tail is eliminated. Capping at 10 yields a more regular distribution: the mean declines to 5.13, the standard deviation to 4.29, and all percentiles above the median collapse to 10. Skewness (0.21) and kurtosis (1.09) indicate an exceptionally well-behaved distribution.

These results demonstrate that both trimming and capping effectively stabilize the weight distributions for both labeling schemes, with capping providing the strongest suppression of extreme observations. Consequently, the main treatment-effect estimates below rely on both trimmed and capped weights as complementary robustness checks.

4.3 RQ1 – Do Type I (third-party certified) eco-labels improve firm financial performance?

Table 4A presents the effects of Type I adoption on firm performance using stabilized IPW estimators. Across all outcomes—Tobin's Q, operating profit (log), operating ratio, and ROA—Type I adoption shows no statistically significant effects under either trimming or capping.

For Tobin's Q, point estimates range from –0.080 (trimmed) to –0.021 (capped), both small in magnitude and statistically indistinguishable from zero. The estimated effects on profitability measures are similarly imprecise and centered near zero. Even after adjusting for selection into certification and stabilizing the weights to address extreme observations, Type I eco-label adoption does not produce measurable financial gains.

Answer to RQ1: Within the study period and after controlling for observable selection, Type I eco-labels do not improve firm financial performance.

4.4 RQ2 – Do Type II (self-declared) eco-labels improve firm financial performance?

Table 4B presents stabilized IPW estimates for Type II adoption. Under 99th-percentile trimming, Type II adoption is associated with a statistically significant increase in operating profit (coefficient = 0.413, $p < 0.01$). Effects on Tobin's Q, operating ratio, and ROA are positive but statistically insignificant. When extreme weights are capped at 10, all estimated effects become statistically insignificant, although coefficients remain positive.

This pattern indicates that gains from Type II labels are sensitive to the treatment of extreme weights. The significant increase in operating profit under trimming does not survive the conservative capping procedure, suggesting the earlier result may be driven by a small number of heavily weighted observations rather than a broad-based performance improvement.

Answer to RQ2: Type II eco-labels show modest and statistically fragile evidence of financial benefits. Any positive effects are not robust across alternative weighting specifications and therefore cannot be interpreted as reliable improvements in firm performance.

Table 4A: Effects of Type I Eco-Label Adoption (IPW with Stabilized Weights)

Outcome	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.
Tobin's Q	2,840	–0.080 (0.065)	–0.021 (0.035)
Operating profit (log)	1,840	0.291 (0.626)	0.155 (0.250)
Operating ratio	2,094	5.887 (8.933)	–0.054 (2.905)
ROA (%)	2,094	0.420 (0.935)	0.139 (0.429)

Notes: Clustered standard errors (in parentheses), clustered at the firm level. All models include firm and fiscal-year fixed effects. Treatment variable: Type I only. **, *** denote significance at the 5%, and 1% levels. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

Table 4B: Effects of Type II Eco-Label Adoption (IPW with Stabilized Weights)

Outcome	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.
Tobin's Q	2,840	0.687	0.065
		(0.581)	(0.086)
Operating profit (log)	1,840	0.413**	0.168
		(0.126)	(0.125)
Operating ratio	2,094	0.477	0.149
		(1.283)	(0.652)
ROA (%)	2,094	0.689	0.320
		(0.529)	(0.402)

Notes: Clustered standard errors (in parentheses), clustered at the firm level. All models include firm and fiscal-year fixed effects. Treatment variable: Type II only. **, *** denote significance at the 5%, and 1% levels. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

4.5 RQ3 – Do the effects of eco-label adoption differ between B2C and B2B firms?

To address RQ3, we classify firms into B2C and B2B categories using NEEDS industry codes. B2C firms operate in sectors where products are sold directly to end consumers (e.g., retail, consumer electronics, food products, personal care). B2B firms include intermediate manufacturers and upstream producers (e.g., industrial equipment, chemicals, components, raw materials) that primarily serve B2B markets. Annually, this corresponds to approximately 175 B2C firms and 393 B2B firms. This classification highlights differences in consumer visibility: eco-labels should be more salient for **B2C** firms, where labels are observed at the point of purchase.

Type I eco-labels by firm type

Table 5 reports heterogeneous treatment effects of Type I adoption by firm type. A clear pattern emerges: Type I labels do not improve financial performance in either group. For B2B firms, all estimated effects are close to zero and statistically insignificant under both trimming and capping. Among B2C firms, coefficients differ in magnitude but remain small and statistically insignificant across all outcomes.

These findings indicate that Type I labels fail to generate financial benefits regardless of a firm's position in the value chain, even in markets with high consumer visibility.

Table 5: Heterogeneous Treatment (Type I) Effects by Firm Type

	B2B Firms			**B2C** Firms		
Outcome	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.
Tobin's Q	1,965	–0.034 (0.044)	–0.028 (0.040)	875	–0.222 (0.223)	–0.020 (0.072)
Operating profit (log)	1,271	0.064 (0.575)	0.017 (0.489)	569	0.247 (0.198)	0.253 (0.182)
Operating ratio	1,432	2.829 (6.490)	0.648 (5.945)	662	-0.449 (1.389)	-0.537 (1.325)
ROA (%)	1,432	0.045 (0.818)	-0.001 (0.708)	662	0.348 (0.545)	0.270 (0.569)

Notes: Clustered standard errors (in parentheses), clustered at the firm level. All models include firm and fiscal-year fixed effects. Treatment variable: Type I only. **, *** denote significance at the 5%, and 1% levels. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

Type II Eco-Labels by firm type

Table 6 presents heterogeneous treatment effects for Type II labels. Among B2B firms, coefficients are positive across all outcomes but do not reach conventional significance levels. This indicates that self-declared labels provide limited informational or reputational value in upstream B2B markets, where buyers do not rely on consumer-facing environmental signals.

Table 6: Heterogeneous Treatment (Type II) Effects by Firm Type

	B2B Firms			**B2C** Firms		
Outcome	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.	Observations	Trim 99th Percentile Coef.	Cap 10 Coef.
Tobin's Q	1,965	0.177 (1.162)	0.108 (0.111)	875	–0.045 (0.041)	–0.102 (0.069)
Operating profit (log)	1,271	0.268 (0.169)	0.152 (0.178)	569	0.552 (0.161)	0.325 (0.166)
Operating ratio	1,432	1.139 (0.729)	0.743 (0.814)	662	0.734 (0.955)	-0.355 (0.840)

ROA (%)	1,432	0.589	0.402	662	1.571	0.592
		(0.462)	(0.522)		(0.505)	(2.000)

Notes: Clustered standard errors (in parentheses), clustered at the firm level. All models include firm and fiscal-year fixed effects. Treatment variable: Type II only. **, *** denote significance at the 5%, and 1% levels. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

In contrast, B2C firms do not exhibit robust financial effects from Type II adoption. Although some coefficients appear positive under the trimmed specification, they are statistically insignificant or highly sensitive to weight stabilization. Once extreme weights are capped, the estimated effects are no longer statistically significant at the conventional 5% level. While the operating-profit (log) coefficient for B2C firms remains positive and close to conventional significance thresholds, it is less robust than under the trimmed specification.

The results suggest that Type II labels may be more significant for consumer-facing B2C firms, where eco-labels are visible at the point of purchase. However, the lack of robustness under capping indicates that the financial relevance of Type II labels in these markets is limited and driven by a subset of firms.

Answer to RQ3: The effects of eco-label adoption do differ by firm type. For B2B firms, neither Type I nor Type II labels generate measurable financial benefits. For B2C firms, Type II labels show weak and specification-sensitive performance gains, while Type I labels remain ineffective.

4.6 Robustness Checks Using Propensity-Score Matching

A potential concern with the IPW–FE estimates is their reliance on reweighting assumptions. Although trimming and capping stabilize the extreme tails of the IPW weights, the estimator may still be sensitive to how reweighting redistributes the effective sample. To ensure that our null results are not artifacts of the weighting procedure, we conduct a robustness check using propensity-score matching (PSM).

PSM avoids reweighting by pairing each treated firm with observationally similar control firms, providing a local comparison estimator based on different identifying assumptions. We implement nearest-neighbor matching with replacement, using a 1-to-1 match and a caliper of 0.05, while imposing the common-support restriction. For this robustness check, propensity scores are estimated using the same core firm-level covariates as in the IPW analysis—debt ratio, employment size, and the Profit margin —augmented with industry and fiscal-year fixed effects. Matching improves covariate balance, with standardized differences substantially reduced relative to the raw sample. The resulting treatment-effect estimates closely align with the IPW findings, confirming that our conclusions do not depend on the weighting procedure.

After constructing nearest-neighbor matched samples with replacement based on the estimated propensity scores, we estimate outcome regressions on the matched sample. Specifically, each performance outcome is regressed on the treatment indicator and the same set of covariates as in the main specifications. The reported treatment effect therefore corresponds to a regression-adjusted matching estimator rather than a simple difference in matched means. Standard errors are clustered at the firm level.

4.6.1 Type I Labels

Table 7 presents the matched results for Type I adopters.
The matching estimates for Type I adoption confirm that Type I eco-labels do not generate measurable financial benefits. Across all four outcome variables—Tobin's Q, log operating profit, operating ratio, and ROA—the estimated treatment effects are small and statistically insignificant. These estimates align with the stabilized-IPW results (Trim99 and Cap10), which also show no statistically significant effect of Type I certification on firm performance. Importantly, the matched estimates use smaller samples due to the loss of observations with poor matches, yet yield the same qualitative conclusion.

After constructing nearest-neighbor matched samples, the reported treatment effects are obtained from outcome regressions estimated on the matched data, controlling for the same covariates as in the main specifications. This reinforces the evidence that the absence of financial effects is not driven by weighting instability or extreme weights.

Table 7: Matching Estimates for Type I Eco-Label Adoption

	(1)	(2)	(3)	(4)
	Tobin's Q	Op.Profit	Op.Ratio	ROA
Type I only	-0.0189	-0.0341	-0.691	-0.308
	(0.0362)	(0.41)	(0.845)	(0.87)
Debt ratio	-0.602	-3.384	-1.175	-6.545
	(0.413)	(2.309)	(3.077)	(4.601)
Employees ×1,000	-0.00166*	0.0179	0.0351**	0.0163***
	(0.0009)	(0.013)	(0.015)	(0.005)
Profit margin	0.131	6.102***	85.55***	16.44***
	(0.213)	(1.504)	(4.300)	(4.599)
Observations	466	434	466	466

Notes: Clustered standard errors (in parentheses), clustered at the firm level. Treatment variable: Type I only. **, *** denote significance at the 5%, and 1% levels. Sample sizes decline because the matching procedure excludes treated firms lacking sufficiently close matches. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

4.6.2 Type II Eco-Label

Table 8 presents the matched results for Type II adopters. The matching estimates for Type II adoption, like those for Type I, provide no evidence that Type II eco-labels improve financial performance. Across all four outcome variables, Tobin's Q, log operating profit, operating-profit ratio, and ROA—the estimated treatment effects are small in magnitude and statistically insignificant.

Although the coefficients are slightly positive, the point estimates are imprecise and economically modest. These findings are fully consistent with the stabilized-IPW estimates (Trim99 and Cap10), which likewise produce no statistically significant evidence of financial gains from adopting a Type II label.

The matching procedure relies on substantially reduced samples due to the exclusion of treated firms lacking close matches; however, the qualitative conclusions remain unchanged. This reinforces the main result: Type II adoption does not yield measurable improvements in firm performance compared to observably similar non-adopters.

Table 8: Matching Estimates for Type II Eco-Label Adoption

	(1)	(2)	(3)	(4)
	Tobin's Q	Op.Profit	Op.Ratio	ROA
Type II only	0.0542	0.308	0.155	0.533
	(0.094)	(0.259)	(0.124)	(0.361)
Debt ratio	-0.596	-2.432*	2.490	-6.844**
	(0.412)	(1.307)	(2.470)	(3.309)
Employees ×1,000	-0.0006	0.0072***	0.0139***	0.0106*
	(0.0004)	(0.0022)	(0.0039)	(0.0055)
Profit margin	0.304	7.990***	88.15***	19.09***
	(0.442)	(2.436)	(5.813)	(4.832)
Observations	420	345	420	420

Notes: Clustered standard errors (in parentheses), clustered at the firm level. Treatment variable: Type II only. **, *** denote significance at the 5%, and 1% levels. Sample sizes decline because the matching procedure excludes treated firms lacking sufficiently close matches. Sample sizes differ across outcomes because each measure requires different accounting variables, and missingness varies across firms and years.

5. Discussion

This study examined whether eco-label adoption improves firms' financial performance using panel data from Japanese listed firms and applying inverse-probability weighting and propensity-score approaches. We found no consistent evidence that either Type I (third-party certified) or Type II (self-declared) labels enhance financial outcomes. This section discusses the emergence of these results from the perspectives of institutional design and market structure, provides answers to the research questions posed in the Introduction, outlines policy implications, and identifies limitations and directions for future research.

5.1. Why do eco-labels fail to improve financial performance? Institutional and market-based explanations

The lack of significant effects for Type I labels can be understood through institutional design and market conditions. While third-party certified labels impose rigorous standards and may enhance firms' environmental practices, the associated costs—such as product redesign, process improvements, and continuous monitoring—are substantial. Prior research shows that formal environmental certification incurs high upfront compliance and verification costs, while efficiency gains accumulate slowly (Darnall & Sides, 2008). Thus, even a five-year observation window may be insufficient for financial returns to emerge.

Market saturation also plays a role. In categories where Type I labels are widely adopted, the marginal signaling value of certification diminishes. When many firms meet the same third-party standard, certification loses its differentiating capacity, limiting its potential to influence consumer demand or generate financial gains (Lyon & Shimshack, 2015; King & Lenox, 2001). Under these conditions, Type I labels may primarily improve internal processes rather than deliver measurable short-run performance effects.

The instability of results for Type II labels reflects the limited informational credibility of self-declared claims. Without external verification, investors and consumers may perceive Type II labels as unreliable or potentially opportunistic indicators of environmental quality. Consequently, Type II labels lack the credibility needed to influence firm performance systematically.

Market structure also explains these findings. In B2B sectors, firms do not interact directly with consumers, resulting in minimal demand-shifting benefits of eco-labels. Environmental preferences are primarily expressed at the final consumer level, resulting in weaker incentives for upstream firms to differentiate themselves through labeling. This structural feature aligns with our heterogeneity results: no meaningful impacts were found for non-B2C firms, while minor patterns appeared only among B2C firms. This finding contrasts with much of the existing firm-level literature, which typically evaluates eco-label adoption without distinguishing firms' positions in the value chain and implicitly assumes uniform financial effects across markets (e.g., Arimura et al., 2011; Schweizer and Zellweger, 2022). By explicitly separating B2C and B2B firms, our analysis shows that the financial relevance of eco-labels is highly context-dependent and largely confined to consumer-facing settings—and even there, the effects are weak and unstable.

5.2. Revisiting the research questions posed in the Introduction

Our findings offer concise answers to the three research questions.

(1) **Do eco-labels improve financial performance?**

No. Neither Type I nor Type II labels generate consistent or economically meaningful financial gains within the institutional and temporal context of this study.

(2) **Do institutional differences between Type I and Type II matter?**

Yes, but not in ways that produce financial benefits. Type I labels show stable null effects across all specifications, and the modest positive estimates for Type II labels disappear under more conservative weighting, indicating a lack of robustness.

(3) **Does market structure condition the effects of eco-labels?**

Yes. Eco-labels have no detectable effects among B2B firms. B2C firms exhibit small, specification-sensitive positive estimates for Type II labels; however, these are not stable enough to support strong conclusions.

Overall, the results confirm the mechanisms proposed in the Introduction: informational asymmetry, signal credibility, and product-market structure jointly determine whether eco-labels can influence firm performance.

5.3. Policy implications

The results suggest that expanding eco-label programs alone is unlikely to generate short-term financial incentives for firms, particularly in upstream or B2B sectors where consumers are not directly involved. Policymakers should therefore differentiate between markets when designing eco-label schemes rather than adopting a "one-size-fits-all" approach.

A useful illustration comes from forest certification systems such as FSC or PEFC, which distinguish between *Forest Management (FM)* certification and *Chain of Custody (CoC)* certification. These programs demonstrate that credible environmental labeling often necessitates monitoring across multiple stages of the production process—not just labeling end products. Similar multi-stage verification mechanisms may be necessary for eco-labels in other sectors, especially where environmental performance depends on supply-chain behavior rather than product attributes alone.

For self-declared Type II claims, the absence of verification underscores the need for stronger oversight, harmonized disclosure guidelines, or third-party auditing to ensure that claims are meaningful and trustworthy. In contrast, for Type I labels, the lack of financial returns despite stringent requirements suggests that adoption may require complementary policies—such as subsidies for certification, tax incentives for sustainable product development, or preferential treatment in public procurement—to offset compliance costs.

Overall, eco-label policies should be integrated into broader environmental governance frameworks that enhance transparency, establish credible information channels throughout the supply chain, and align incentives for both upstream and downstream actors.

5.4. Limitations and directions for future research

This study has several limitations that suggest avenues for further inquiry. First, the observation period (2012–2016) may be too short to capture the long-term financial implications of eco-label adoption, particularly for Type I certifications that require sustained investments. Longer panels could reveal dynamic effects not observable in the short run.

Second, the analysis focuses on Type I and Type II labels due to data availability. Other schemes—such as Type III environmental product declarations or international certifications (e.g., EU Ecolabel, Energy Star)—may operate through different mechanisms and merit comparative analysis.

6. Conclusion

This study investigated whether eco-label adoption improves firm-level financial performance using panel data on Japanese listed firms from 2012 to 2016. By estimating separate propensity score models for Type I (third-party certified) and Type II (self-declared) labels and applying inverse probability weighting with firm and year fixed effects, we addressed observable selection into certification. Due to the extreme heavy-tailedness of raw IPW weights, we employed two stabilization methods—99th-percentile trimming and capping at 10—to ensure robust inference.

Across all specifications, we find no consistent evidence that eco-labels enhance firm performance. For Type I labels, all outcome measures are statistically insignificant under both trimming and capping. For Type II labels, trimming shows a positive association with operating profit and ROA in certain models, but these effects disappear under capped weights, indicating instability. The heterogeneity analysis reinforces these conclusions: B2B firms primarily in upstream or B2B markets show no measurable performance improvements from either Type I or Type II labels. In contrast, B2C firms exhibit modest gains from Type II labels under trimmed weights, but these effects dissipate under capped weights, indicating fragile benefits that are context-dependent. Overall, these findings challenge the notion that eco-labels deliver “win–win” benefits for firms and the environment.

For policymakers, these results suggest that eco-label programs should be justified based on their environmental effectiveness rather than on presumed financial incentives. If financial returns are limited or inconsistent, relying solely on market-based incentives may not be sufficient to encourage the widespread adoption of credible eco-labels.

For firms, the evidence suggests that certification alone is unlikely to yield substantial operational or market-based returns. Eco-labels should be pursued primarily for non-financial reasons—such as regulatory compliance, stakeholder relations, risk management, or genuine environmental commitment—rather than expectations of immediate profitability gains.

Future research should examine whether eco-labels generate environmental benefits without financial returns, whether financial effects differ in less saturated or emerging markets, and whether alternative identification strategies—such as natural experiments or instrumental-variable approaches—reveal causal mechanisms that our IPW–FE framework may not fully capture. Addressing these questions will provide a more comprehensive understanding of when and how eco-labels contribute to environmental and organizational outcomes.

References

-Abadie, A., & Imbens, G. (2016). *Matching on the estimated propensity score*. Econometrica, 84(2), 781–807. DOI: https://doi.org/10.3982/ECTA11293

-Aghion, P., Bloom, N., Blundell, R., Griffith, R., & Howitt, P. (2005). *Competition and innovation: An inverted-U relationship*. The Quarterly Journal of Economics, 120(2), 701–728. DOI: https://doi.org/10.1093/qje/120.2.701

-Arimura, T., Darnall, N., & Katayama, H. (2011). *Is ISO 14001 a gateway to more advanced voluntary action?* JEEM, 61(2), 170–182. DOI: https://doi.org/10.1016/j.jeem.2010.11.003

-Austin, P. C., & Stuart, E. A. (2015). Moving towards best practice when using inverse probability weights in observational studies. Statistics in Medicine, 34(28), 3661–3679. DOI: 10.1002/sim.6607

-Brécard, D., Hlaimi, B., Lucas, S., Perraudin, C., & Salladarré, F. (2009). *Determinants of demand for green products: An application to eco-label demand for fish in Europe*. Ecological Economics, 69(1), 115–125. DOI: https://doi.org/10.1016/j.ecolecon.2009.07.017

-Busso, M., DiNardo, J., & McCrary, J. (2014). *New evidence on the finite-sample properties of propensity score reweighting and matching estimators*. Review of Economics and Statistics, 96(5), 885–897. DOI: https://www.jstor.org/stable/43554965

-Cason, T. N., & Gangadharan, L. (2002). *Environmental labeling and incomplete consumer information in laboratory markets*. Journal of Environmental Economics and Management, 43(1), 113–134. DOI: https://doi.org/10.1006/jeem.2000.1170

-Cole, S. R., & Hernán, M. A. (2008). Constructing inverse probability weights for marginal structural models. Epidemiology, 19(3), 278–284. DOI: https://doi.org/10.1097/EDE.0b013e3181640d76

-Consumers International (2022). *The State of Sustainability Information: Critical Trends, Trade-offs, and Solutions.* [PDF] Available at: https://www.consumersinternational.org/media/451292/the-state-of-sustainability-information.pdf

-Crump, R., Hotz, V., Imbens, G., & Mitnik, O. (2009). Dealing with limited overlap in estimation of average treatment effects. Econometrica, 77(6), 1711–1745. DOI: https://doi.org/10.3982/ECTA7220

-Darnall, N., & Sides, S. (2008). *Assessing the performance of voluntary environmental programs: Does certification matter?* Policy Studies Journal, 36(1), 95–117. DOI: https://doi.org/10.1111/j.1541-0072.2007.00255.x

-Greenstone, M., Gayer, T., & Li, S. (2012). *Quasi-experimental and experimental approaches to environmental economics*. Journal of Environmental Economics and Management, 64(1), 1–3. DOI: https://doi.org/10.1016/j.jeem.2008.02.004

-King, A., & Lenox, M. (2001). *Does it really pay to be green? An empirical study of firm environmental and financial performance*. Journal of Industrial Ecology, 5(1), 105–116. DOI: https://doi.org/10.1162/108819801753358526

-Lyon, T. P., & Shimshack, J. P. (2015). *Environmental disclosure: Evidence from economics and* *policy*. Foundations and Trends in Microeconomics, 11(3), 153–307. DOI: https://doi.org/10.1561/0700000050

-Meis-Harris, J., Narasimhan, A., & Coady, D. (2021). *Global eco-labels database*. Ecolabel Index.

-Nimon, W., & Beghin, J. (1999). *Are eco-labels valuable? Evidence from the apparel industry*. American Journal of Agricultural Economics, 81(4), 801–811. DOI: https://doi.org/10.2307/1244325

-OECD (2016). *Environmental labelling and information schemes: Policy perspectives*. OECD Publishing.

-Schleenbecker, R., & Hamm, U. (2013). *Consumers' perception of organic product characteristics: A review*. Renewable Agriculture and Food Systems, 28(4), 344–351. DOI: https://doi.org/10.1017/S174217051200058X

-Schweizer, M., & Zellweger, T. (2022). *Do eco-labels create value? Firm-level evidence from international markets*. Business Strategy and the Environment, 31(5), 1–15. DOI: https://doi.org/10.1002/bse.3012

-Spence, M. (1973). *Job market signaling*. Quarterly Journal of Economics, 87(3), 355–374. DOI: https://doi.org/10.2307/1882010

-Teisl, M. F., Rubin, J., & Noblet, C. (2008). *Non-dirty dancing? Eco-labels and market behavior*. In *Frontiers in Environmental Economics* (pp. 123–146). Edward Elgar Publishing. DOI: https://doi.org/10.1016/j.joep.2007.04.002

-UN Environment Programme (2018). *Recommendations for Eco-Labelling Platform for Sri Lanka*. Colombo: Ministry of Mahaweli Development and Environment.
Retrieved from https://www.scp.env.gov.lk/kdb/docs/Report%20Eco%20Labeling%20Platform%2020180511.pdf

Appendix

Table A-1. Pairwise Mean Comparisons (Bonferroni-adjusted p-values)

Panel A. Debt ratio

Comparison	Mean difference	Std. Error	95% CI	Significant?
Type I vs Non	–0.0039	0.0087	[–0.0246, 0.0168]	No
Type II vs Non	–0.0124	0.0094	[–0.0349, 0.0101]	No
Type II vs Type I	–0.0085	0.0118	[–0.0369, 0.0198]	No

Panel B. Employees (×1,000)

Comparison	Mean difference	Std. Error	95% CI	Significant?
Type I vs Non	7.513	1.799	[3.203, 11.823]	Yes
Type II vs Non	13.821	1.949	[9.153, 18.489]	Yes
Type II vs Type I	6.308	2.446	[0.447, 12.169]	Yes

Panel C. Profit margin

Comparison	Mean difference	Std. Error	95% CI	Significant?
Type I vs Non	0.0544	0.0118	[0.0261, 0.0826]	Yes
Type II vs Non	0.0112	0.0126	[–0.0190, 0.0414]	No
Type II vs Type I	–0.0432	0.016	[–0.0816, –0.0048]	Yes

Panel D. Sales (million yen)

Comparison	Mean difference	Std. Error	95% CI	Significant?

Type I vs Non	313,845	44,293	[207,726, 419,964]	Yes
Type II vs Non	103,520	47,650	[–10,643, 217,683]	No
Type II vs Type I	–210,325	60,400	[–355,036, –65,615]	Yes

Panel E. Tobin's Q

Comparison	Mean difference	Std. Error	95% CI	Significant?
Type I vs Non	0.0118	0.0465	[–0.0997, 0.1233]	No
Type II vs Non	0.0724	0.0506	[–0.0487, 0.1935]	No
Type II vs Type I	0.0606	0.0637	[–0.0919, 0.2131]	No

Panel F. Operating profit (million yen)

Comparison	Mean difference	Std. Error	95% CI	Significant?
Type I vs Non	22,517	4,122	[12,642, 32,393]	Yes
Type II vs Non	12,800	4,407	[2,240, 23,359]	Yes
Type II vs Type I	–9,717	5,605	[–23,145, 3,710]	No

Panel G. ROA (%)

Comparison	Mean difference	Std. Error	95% CI	Significant?

Type I vs Non	–0.063	0.297	[–0.774, 0.648]	No
Type II vs Non	–0.654	0.317	[–1.414, 0.106]	No
Type II vs Type I	–0.591	0.403	[–1.557, 0.376]	No