Transboundary Haze and Air Pollution in Southeast Asia: A Game-Theoretic and Panel-Data Analysis

Simon Poltak Hamonangan Hutabarat, Ph.D.*[1,2] and Irfan Yulianto, S.E., M.M.†[1]

[1]Directorate General of Taxes, Ministry of Finance, Republic of Indonesia
[2]Faculty of Business, Sampoerna University, Indonesia

Abstract

Transboundary haze pollution, driven primarily by land-clearing fires in Indonesia, has recurred across Southeast Asia for over two decades despite repeated regional policy commitments. This paper examines the persistence of this problem using two complementary approaches. First, we develop a game-theoretic model of transboundary externalities in which a source country's privately optimal burning intensity exceeds the socially efficient level, and derive the condition under which a compensation mechanism – modeled on the ASEAN Transboundary Haze Pollution Control Fund – can sustain cooperation in an infinitely repeated game. Second, we test the model's core empirical prediction using a novel panel dataset combining NASA satellite fire-detection data (2015–2024) with World Bank economic and environmental indicators for all ten ASEAN member states, employing a shift-share instrumental-variable strategy based on the El Niño Southern Oscillation. We find that fire exposure significantly increases PM2.5 concentrations among transboundary-recipient countries ($\beta \approx 0.017$, $p < 0.01$), a result that is robust to excluding the COVID-19 period, to instrumental-variable estimation (first-stage $F = 82.1$), and that is statistically indistinguishable from zero in a domestic-pollution placebo sample. The corresponding health effect could not be estimated with comparable precision, since World Bank mortality data are available for only one cross-sectional year in our sample window; we report this data limitation explicitly. We further identify an unresolved theoretical tension: our baseline calibration implies cooperation should be sustainable at modest compensation-fund levels, a prediction difficult to reconcile with over two decades of recurring non-compliance, suggesting that collective-action and enforcement-credibility problems, rather than fund size, are the binding constraint on regional cooperation.

Keywords: transboundary pollution; game theory; PM2.5; ASEAN; instrumental variables; shift-share design; panel data

JEL Classification: Q53, Q58, C73, F53, I18

*Corresponding author. Emails: simon.hutabarat@kemenkeu.go.id; simon.hutabarat@sampoernauniversity.ac.id

†Email: irfan.yulianto@kemenkeu.go.id. The views expressed in this paper are solely those of the authors and do not necessarily represent the official position of the Directorate General of Taxes or the Ministry of Finance of the Republic of Indonesia.

1 Introduction

Transboundary haze pollution, generated primarily by land-clearing and peatland fires in Indonesia, has affected Southeast Asia intermittently since at least 1997, with major episodes recurring in 2006, 2013, 2015, and 2019 (World Bank, 2016; Koplitz et al., 2016). As this manuscript was being prepared, the Singapore Institute of International Affairs issued a "red," or highest-risk, transboundary haze outlook for 2026 – only the second such designation since the outlook series began in 2019 – citing a strong El Niño Southern Oscillation (ENSO) event. The recurrence of severe haze episodes despite nearly a quarter-century of regional policy attention, most notably the 2002 ASEAN Agreement on Transboundary Haze Pollution (AATHP), motivates the two questions addressed in this paper: (i) under what conditions can a compensation mechanism resolve the underlying transboundary externality, and (ii) does the region's satellite and economic data support the causal chain – from fire activity to air pollution to, ultimately, adverse health outcomes – that the policy debate presumes.

The economic magnitude of the problem is well established in the existing literature. The 2015 fire season alone has been estimated to have cost the Indonesian economy between US$16.1 billion, or approximately 1.9 percent of GDP (World Bank, 2016), and US$28 billion in a subsequent peer-reviewed estimate of avoided losses from peatland restoration (Castro et al., 2026), and to have caused approximately 100,300 excess deaths across Indonesia, Malaysia, and Singapore combined (Koplitz et al., 2016). An earlier and widely cited study attributes approximately 15,600 excess infant deaths in Indonesia to the smaller 1997 fire episode (Jayachandran, 2009).

1.1 Research Gap

Despite this well-documented health and economic burden, the existing literature on Southeast Asian haze exhibits a persistent disconnect between two research traditions. A substantial body of political-science and legal scholarship analyzes the institutional weaknesses of the AATHP, generally attributing its limited effectiveness to ASEAN's consensus-based, non-interference norms (Muhammad, 2022; Chaves et al., 2025). A separate empirical literature, most notably Sheldon and Sankaran (2017), uses satellite fire data as an instrument to estimate the causal effect of Indonesian fires on pollution and health outcomes in specific downwind locations, typically Singapore. To our knowledge, no existing study combines a formal game-theoretic treatment of the AATHP's compensation mechanism, calibrated to its documented funding history, with a panel test of the mechanism's underlying empirical prediction across the full set of ASEAN member states. This paper is intended to address that gap, while remaining explicit about the limitations such an integration necessarily entails.

1.2 Contribution and Summary of Main Findings

This paper makes three contributions. First, we formalize the transboundary haze externality as a repeated game and derive, in closed form, the discount-factor condition under which a compensation fund modeled on the actual AATHP mechanism can sustain cooperation (Section 3.1; full derivation in Appendix A). Second, we construct a shift-share (Bartik-style) instrumental variable, following the methodological framework of Goldsmith-Pinkham et al. (2020), that permits identification of the fire-pollution relationship in a two-way fixed-effects panel – a design choice that is not merely convenient but, as we demonstrate in Section 3.4, necessary given

that raw fire counts and the ENSO index are collinear with time fixed effects. Third, we test this design on a newly assembled ten-country, ten-year panel (2015–2024) and report both the well-identified pollution-pathway result and the comparatively underpowered mortality result with equal candor.

Our principal empirical finding is that fire exposure raises PM2.5 concentrations among transboundary-recipient countries by approximately 0.017 $\mu g/m^3$ per unit of exposure ($p < 0.01$), a result robust to instrumental-variable estimation and absent in a domestic-pollution placebo sample. We do not find a statistically distinguishable effect on mortality, a result we attribute to data availability rather than to the absence of a true effect. Finally, we document a theoretical tension between our model's prediction that cooperation should be readily sustainable and the empirical persistence of non-compliance, which we interpret, following the discussion in Section 5.3, as evidence for a collective-action problem not captured by a single-agent representation of the source country.

1.3 Organization of the Paper

Section 2 reviews the relevant literature and situates this paper's contribution within it. Section 3 presents the theoretical model, data, and empirical strategy. Section 4 reports results. Section 5 discusses these results in relation to the existing literature and addresses the study's limitations. Section 6 concludes with policy recommendations.

2 Literature Review

2.1 Transboundary Pollution and Repeated Games

The theoretical treatment of transboundary pollution as a negative externality dates to the classical environmental economics literature (Pigou, 2017; Baumol and Oates, 1988), which established that, absent corrective intervention, a polluter who does not internalize downwind damages will select a privately optimal emission level exceeding the socially efficient level. The standard remedies proposed in this literature are Coasian side-payments (Coase, 2013) and repeated-game cooperation sustained by the threat of retaliation (Barrett, 2001). Applied specifically to transboundary air pollution, Yu et al. (2024) model cooperative and non-cooperative control strategies between regions of asymmetric economic development in a differential-game framework, finding that a cooperative arrangement paired with a compensation mechanism for the less-developed region dominates unilateral action for both parties. Our theoretical model in Section 3.1 adopts a similar repeated-game structure but calibrates the compensation mechanism specifically to the documented funding history of the AATHP (Section 2.3), rather than treating fund size as a free parameter.

2.2 Empirical Evidence on Southeast Asian Haze

The empirical literature on Indonesian fires and downwind outcomes is comparatively well developed. Sheldon and Sankaran (2017) provide what is, to our knowledge, the first causal estimate of the effect of Indonesian forest fires on a downwind country's pollution and health outcomes, instrumenting Singaporean air pollution with satellite-derived fire radiative power and finding that a one-standard-deviation increase in fire intensity raises Singapore's Pollution Standards

Index by approximately 16 percent and increases polyclinic attendance for acute respiratory infection and conjunctivitis by more than 1.5 percent. A companion study documents defensive behavioral responses among Singaporean households during haze episodes (Sheldon and Sankaran, 2019). On health outcomes specifically, Koplitz et al. (2016) use chemical transport modeling to estimate approximately 100,300 excess deaths across Indonesia, Malaysia, and Singapore attributable to the September–October 2015 haze episode, extending an earlier estimation framework applied to the 2004–2015 period (Kiely et al., 2020). Jayachandran (2009) provides an earlier and methodologically distinct contribution, using discrepancies between expected and observed birth cohorts in Indonesia's 2000 census to estimate approximately 15,600 excess infant deaths attributable to the 1997 fires. An earlier study of forest fires, air pollution, and mortality across Southeast Asia established the demographic methodology on which subsequent work, including Jayachandran (2009), builds (Sastry, 2002).

2.3 Institutional Analysis of the AATHP

A distinct literature examines the AATHP's institutional design and implementation record. This literature generally attributes the agreement's limited enforcement capacity to ASEAN's foundational norm of non-interference in member states' internal affairs, which is argued to be structurally incompatible with the binding obligations and third-party dispute resolution that credible international environmental agreements typically require (Muhammad, 2022). A related legal analysis situates the AATHP within the broader supranational framework for the right to a healthy environment in Southeast Asia, reaching a similar conclusion regarding weak enforceability (Chaves et al., 2025). A more recent institutionalist account characterizes ASEAN's progression from the 1997 Regional Haze Action Plan through the 2002 AATHP to the 2024 "Second Haze-Free Roadmap" not as institutional stagnation but as a deliberate, if limited, adaptation to the region's political constraints (Ryu and Lee, 2026). Consistent with this interpretation, Pham and Thomson (2025) document that ASEAN member states systematically delegate only limited discretionary power in their international agreements more broadly, a pattern that helps explain the AATHP's specific design choices rather than treating them as anomalous. A related study documents the recurrence of haze crises since 2023 and the continued absence of enforceable regional commitments (Putra, 2024).

2.4 Research Gap and Contribution of This Study

Sections 2.1–2.3 illustrate a gap this paper aims to address: the game-theoretic literature on transboundary pollution (Section 2.1) is rarely calibrated to a specific, empirically documented institutional mechanism, the empirical literature on Southeast Asian haze (Section 2.2) is generally confined to a single country or fire episode, and the institutional literature (Section 2.3) explains the AATHP's weakness primarily through political rather than economic logic. This paper contributes to closing this gap along three dimensions. First, our theoretical model (Section 3.1) is calibrated directly to the AATHP's documented funding shortfall, yielding a quantitative, testable prediction regarding the fund size required to sustain cooperation. Second, our empirical design (Sections 3.2–3.4) extends the single-country satellite-instrument approach of Sheldon and Sankaran (2017) to a ten-country panel, allowing us to test the fire-pollution relationship across the full set of transboundary-recipient states while using a domestic-pollution comparison group as an internal placebo. Third, we explicitly reconcile our theoretical and em-

pirical findings in Section 5.3, arguing that the divergence between the model's prediction and the observed persistence of non-compliance is itself informative about the AATHP's institutional design – a synthesis we do not find elsewhere in the literature reviewed above.

3 Materials and Methods

3.1 Theoretical Framework

We model the transboundary haze problem as a repeated game between a source country S (representing Indonesia's land-clearing sector in aggregate) and a representative downwind receiver country R (representing the transboundary-recipient states defined in Section 3.2).

3.1.1 Static Stage Game

In each period, the source chooses a burning or land-clearing intensity $e_S \geq 0$, empirically proxied in Section 3.2 by satellite-detected fire activity. Burning intensity generates a private benefit – the value of cleared agricultural land – net of a private cost, giving the stage payoff

$$\pi_S(e_S) = a\, e_S - \frac{b}{2} e_S^2 - pF \cdot \mathbb{1}\{e_S > \tau\} \tag{1}$$

where $a > 0$ is the marginal private benefit of burning at $e_S = 0$, $b > 0$ governs the rate at which marginal benefit declines, and the final term represents expected enforcement: p is the probability that AATHP-style enforcement is triggered, F is the associated penalty, and τ is a compliance threshold. Historically, $p \approx 0$ under the AATHP's voluntary compliance regime (Section 2.3), so we set $p = 0$ in the baseline case and treat $p > 0$ as a comparative-statics exercise (Appendix A).

The receiver country suffers linear damage $D(e_S) = d\, e_S$ from the source's burning, where $d > 0$ is the marginal transboundary damage per unit of burning intensity, and may undertake adaptive investment a_R (e.g., air purification, school closures) at convex cost, yielding

$$\pi_R = -D(e_S)(1 - a_R) - \frac{k}{2} a_R^2 \tag{2}$$

where $k > 0$ parameterizes the cost of adaptation. Because the receiver's adaptation decision does not affect the source's incentives in this model – the source does not internalize $D(e_S)$ regardless of a_R – the receiver's problem does not enter the source's first-order condition and is presented here for completeness rather than solved explicitly.

With $p = 0$, the source's non-cooperative (Nash) choice of burning intensity solves $\max_{e_S} \pi_S(e_S)$. The first-order condition is

$$\frac{\partial \pi_S}{\partial e_S} = a - b\, e_S = 0 \quad \Longrightarrow \quad e_S^N = \frac{a}{b} \tag{3}$$

A social planner maximizing joint surplus $\pi_S + \pi_R$ with respect to e_S – holding a_R at its optimum, which does not affect the first-order condition in e_S – instead solves

$$\frac{\partial (\pi_S - D)}{\partial e_S} = a - b\, e_S - d = 0 \quad \Longrightarrow \quad e_S^* = \frac{a - d}{b} \tag{4}$$

Comparing (3) and (4) yields the paper's first testable prediction: $e_S^N > e_S^*$ whenever $d > 0$, i.e., the source burns strictly more than is socially optimal whenever transboundary damage is positive, and the magnitude of over-burning, $e_S^N - e_S^* = d/b$, is increasing in the severity of downwind damage. Section 4.1 reports this gap under an illustrative calibration; Appendix A.1 derives the associated deadweight-loss expression.

3.1.2 Repeated Game with a Compensation Mechanism

We model the AATHP's compensation fund as supporting a grim-trigger strategy in an infinitely repeated version of the stage game. Under this strategy, the receiver (and, by extension, the wider ASEAN membership) sustains cooperation – the source plays e_S^* – as long as no deviation has occurred, and permanently withdraws fund access and cooperative support following any deviation, after which the source receives a fixed per-period punishment payoff π_S^{punish} for the remainder of the game. We model π_S^{punish} as decreasing in fund size F, reflecting the intuition that a larger fund represents a larger forfeited benefit upon punishment: $\pi_S^{\text{punish}} = -\kappa F$ for a scaling parameter $\kappa > 0$ ($\kappa = 0.5$ in our baseline calibration, Section 4.1).

Let $\delta \in (0, 1)$ denote the source's discount factor. Cooperation is sustainable in a subgame-perfect equilibrium if and only if the value of perpetual cooperation weakly exceeds the value of one-shot deviation followed by permanent punishment, which – as derived in full in Appendix A.2 – reduces to

$$\delta \geq \delta^*(F) \equiv \frac{\pi_S(e_S^N) - \pi_S(e_S^*)}{\pi_S(e_S^N) - \pi_S^{\text{punish}}(F)} \tag{5}$$

Equation (5) is the paper's second testable prediction: $\delta^*(F)$ is decreasing in F (a larger fund lowers the patience required to sustain cooperation), and Section 4.1 reports its numerical behavior under our baseline calibration, together with the comparative statics with respect to the damage parameter d (Appendix A.3).

3.2 Data Sources and Variable Construction

We assemble a country-year panel covering all ten ASEAN member states (Indonesia, Malaysia, Singapore, Brunei, Thailand, Vietnam, the Philippines, Myanmar, Laos, and Cambodia) from five sources. NASA's Fire Information for Resource Management System (FIRMS) provides daily active-fire detections from the MODIS and VIIRS-SNPP satellite sensors for 2015–2024; detections are restricted to the "vegetation fire" classification, excluding active volcanoes and gas flares, both of which are otherwise present in the raw detection stream and are particularly relevant for Indonesia. World Bank Open Data provides annual PM2.5 (mean population-weighted exposure, indicator `EN.ATM.PM25.MC.M3`), GDP per capita (`NY.GDP.PCAP.KD`), population (`SP.POP.TOTL`), health expenditure as a share of GDP (`SH.XPD.CHEX.GD.ZS`), and the mortality rate attributable to air pollution (`SH.STA.AIRP.P5`). NOAA's Oceanic Niño Index (ONI), averaged over the June–November fire season, provides the instrument described in Section 3.4.

Countries are classified into three channels based on their primary exposure pathway: transboundary recipients (Malaysia, Singapore, Brunei, Thailand), for whom Indonesian fires are the dominant hypothesized pollution channel; a domestic-channel comparison group (Vietnam, Philippines), whose pollution is attributed primarily to vehicular and industrial sources unrelated to Indonesian fires; and source countries (Indonesia, Myanmar, Laos, Cambodia), whose

own fire activity is the explanatory variable rather than an outcome in this study.

The central explanatory variable, fire exposure, is constructed as

$$\text{FireExposure}_{it} = w_i \times \text{Fires}_{\text{IDN},t} \tag{6}$$

where $\text{Fires}_{\text{IDN},t}$ is Indonesia's national vegetation-fire count in year t and w_i is a country-specific exposure weight (1.3 for Malaysia, 1.0 for Singapore, 0.7 for Thailand, 0.6 for Brunei, 0.08 for Vietnam, 0.03 for the Philippines, and 0 for the four source countries). We emphasize that these weights are illustrative, reflecting approximate geographic plausibility rather than an estimated quantity; we return to this limitation in Section 5.5. The corresponding instrument, described formally in Section 3.4, is constructed analogously as $\text{OniExposure}_{it} = w_i \times \text{ONI}_t$.

The principal data limitation in this study concerns mortality: World Bank indicator `SH.STA.AIRP.P5`, a modeled WHO estimate, is non-missing only for 2019 in our extraction window, with one observation per country. This constrains the mortality analysis in Section 4.3 to a ten-observation cross-section rather than a panel, a limitation we address directly in the empirical strategy (Section 3.4.2) and revisit in the discussion (Section 5.2).

3.3 Descriptive Statistics

Table 1: Summary statistics

Group	Variable	Mean	SD	Min	Max	N
All ASEAN-10 (panel)	PM2.5 (μg/m^3)	21.75	7.20	6.42	36.88	90
All ASEAN-10 (panel)	Fire hotspots (thousands)	118.39	149.00	0.00	859.41	100
All ASEAN-10 (panel)	Fire exposure	81.25	176.40	0.00	1117.23	100
All ASEAN-10 (panel)	GDP per capita (US$)	12475.71	18599.54	1136.996	68271.09	100
All ASEAN-10 (panel)	Health exp. (% GDP)	3.91	1.07	1.33	6.34	90
Channel = recipient	PM2.5 (μg/m^3)	16.51	6.29	6.42	26.79	36
Channel = recipient	Fire hotspots (thousands)	45.01	79.75	0.00	235.62	40
Channel = recipient	Fire exposure	197.11	236.38	5.33	1117.23	40
Channel = recipient	GDP per capita (US$)	27172.08	22530.17	5688.85	68271.09	40
Channel = recipient	Health exp. (% GDP)	3.76	0.99	1.77	5.70	36
Channel = domestic	PM2.5 (μg/m^3)	24.39	2.03	21.77	27.47	18
Channel = domestic	Fire hotspots (thousands)	49.78	28.38	2.07	107.71	20
Channel = domestic	Fire exposure	12.05	15.89	0.27	68.75	20
Channel = domestic	GDP per capita (US$)	3326.66	391.41	2577.57	4012.51	20
Channel = domestic	Health exp. (% GDP)	4.66	0.56	3.90	5.98	18
Channel = source	PM2.5 (μg/m^3)	25.66	6.53	17.77	36.88	36
Channel = source	Fire hotspots (thousands)	226.07	172.55	8.88	859.41	40
Channel = source	Fire exposure	0.00	0.00	0.00	0.00	40
Channel = source	GDP per capita (US$)	2353.88	976.94	1136.996	4367.95	40
Channel = source	Health exp. (% GDP)	3.68	1.19	1.33	6.34	36
2019 cross-section only	Mortality (per 100k)	111.04	70.93	19.60	202.80	10

Table 1 reports means, standard deviations, and ranges for the core panel variables, overall and by channel group, with the 2019 mortality cross-section reported separately given its distinct sample. Transboundary-recipient countries exhibit the lowest mean PM2.5 (16.51 μg/m^3) of the three groups, reflecting the inclusion of Singapore and Brunei, both comparatively low-pollution states outside acute haze episodes. Source countries exhibit both the highest mean

PM2.5 (25.66 μg/m^3) and the lowest mean income (US$2,354 per capita, versus US$27,172 for the recipient group), consistent with the well-documented association between fire-based land clearing and lower-income agricultural economies. The domestic-channel group exhibits the smallest standard deviation in PM2.5 (2.03) of the three groups, consistent with a pollution process driven by a comparatively stable urbanization trend rather than the episodic variation characteristic of fire-driven haze. Mean fire hotspot counts follow the expected ordering across channel groups (source: 226.07 thousand; domestic: 49.78 thousand; recipient: 45.01 thousand), reflecting each group's differing exposure to and, for source countries, direct contribution to regional fire activity; health expenditure as a share of GDP is highest among domestic-channel countries (4.66 percent) and comparable between the recipient and source groups (3.76 and 3.68 percent, respectively).

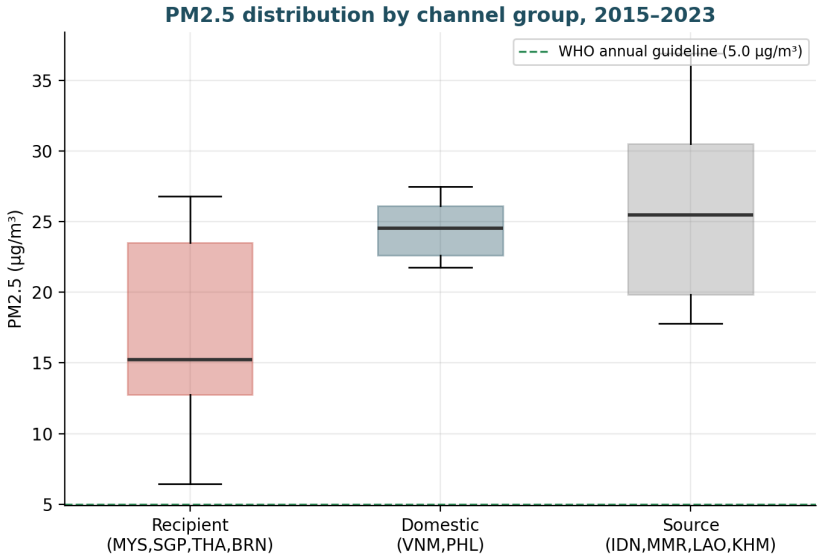


Figure 1: Distribution of PM2.5 by channel group, 2015–2023.

Figure 1 presents the channel-group distributions in Table 1 as box plots. The recipient group (red) exhibits both the lowest median and the widest interquartile range of the three groups, consistent with a pollution process subject to episodic fire-driven spikes superimposed on a lower baseline. All three group distributions lie substantially above the WHO annual PM2.5 guideline of 5 μg/m^3, indicated by the dashed reference line.

Figure 2 disaggregates Figure 1 into individual country trajectories. Source-country trends (grey), particularly Myanmar, remain persistently elevated with comparatively little year-to-year variation. Several recipient-country trends (red), notably Malaysia and Singapore, exhibit visible peaks in 2015 and 2019, corresponding to documented severe haze episodes (Section 2.2).

Figure 3 ranks countries by mean PM2.5. Channel-group membership (indicated by bar color) does not correspond to a monotonic ranking: recipient-group countries (red) appear at both the cleanest (Brunei) and mid-range (Thailand) positions, interspersed with source and domestic-channel countries, indicating that mean pollution levels reflect multiple determinants – income, urbanization, geography – of which transboundary haze exposure is one among several.

Figure 4 plots the bivariate relationship between fire exposure and PM2.5 across all country-year observations, with a fitted line restricted to the recipient subsample. The fitted relationship

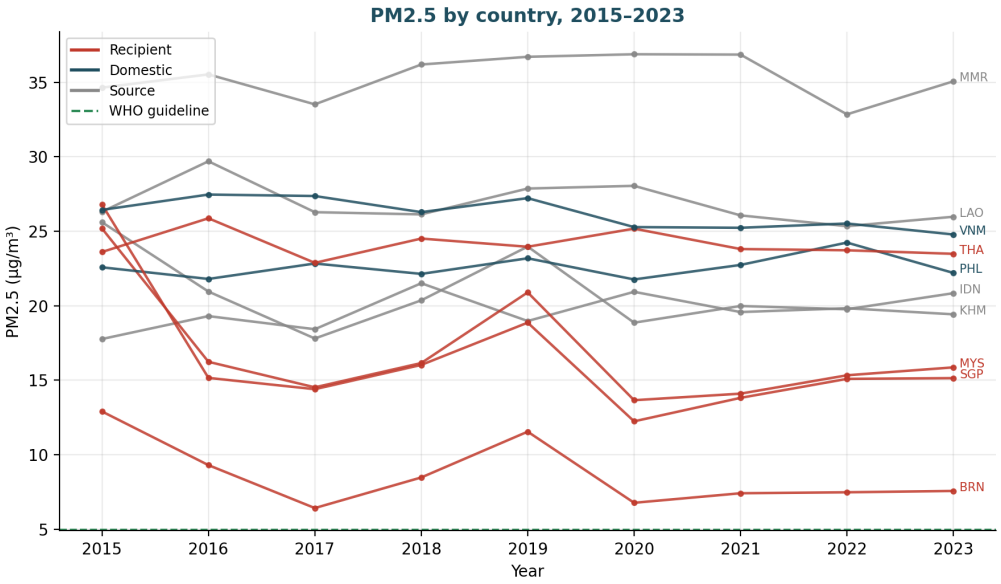


Figure 2: Country-level PM2.5 trends, 2015–2023.

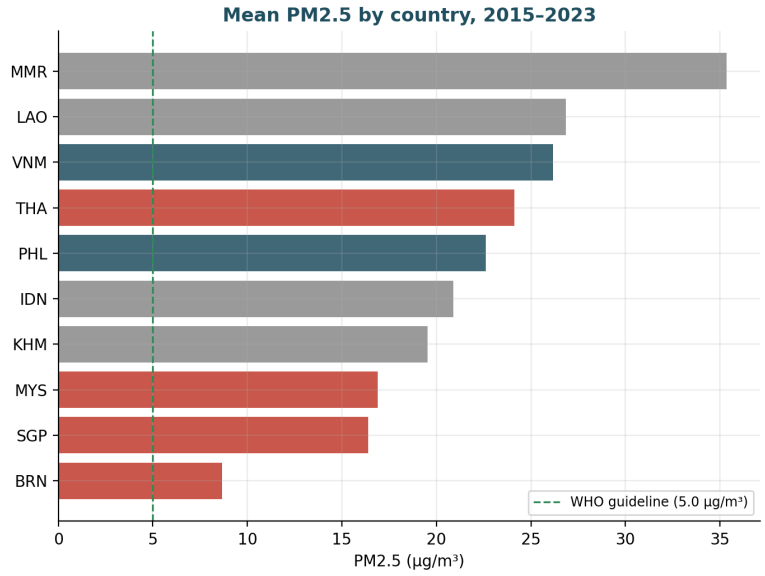


Figure 3: Mean PM2.5 by country, 2015–2023, ranked against the WHO guideline.

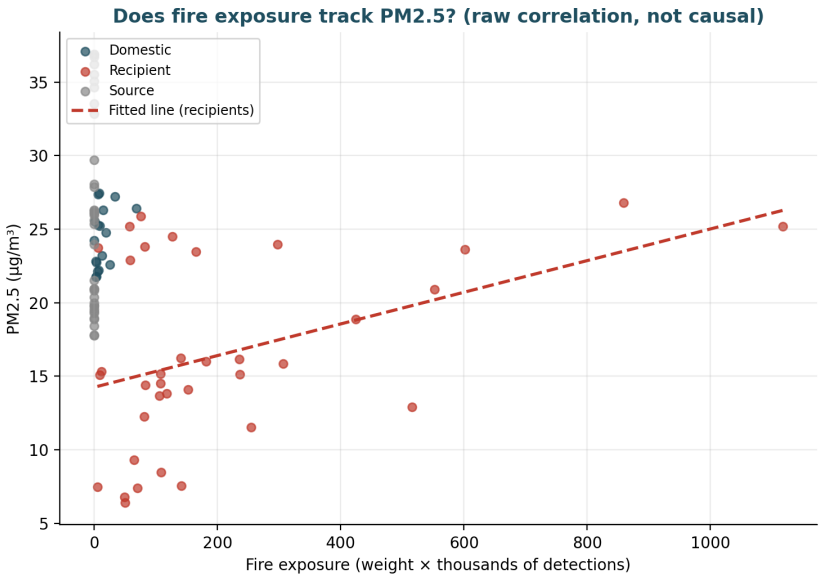


Figure 4: Bivariate relationship between fire exposure and PM2.5, by channel.

is positive, consistent with the hypothesized mechanism; we note explicitly that this bivariate relationship is descriptive and does not account for confounding trends common to fire activity and economic development, which the identification strategy in Section 3.4 is designed to address.

Figure 5 reports pairwise correlations among the key panel variables. Fire exposure correlates with PM2.5 at 0.44 and with the ONI-based instrument at 0.73, consistent with the first-stage relationship exploited in Section 3.4. Log GDP per capita correlates strongly with mortality (-0.84), motivating its inclusion as a control variable throughout the empirical specifications.

3.4 Empirical Strategy

3.4.1 Pollution Pathway (Panel Estimation)

We estimate the effect of fire exposure on PM2.5 using a two-way fixed-effects specification:

$$\text{PM2.5}_{it} = \beta_1 \cdot \text{FireExposure}_{it} + \beta_2 \cdot X_{it} + \alpha_i + \gamma_t + \varepsilon_{it} \tag{7}$$

where α_i and γ_t are country and year fixed effects and X_{it} includes log GDP per capita and health expenditure as a share of GDP. As noted in Section 3.2, we deliberately use the exposure-weighted variable FireExposure_{it} rather than the raw national fire count: because the raw count and the raw ONI index are identical across countries within a year, either variable is exactly collinear with γ_t once year fixed effects are included, and the shift-share interaction with w_i is what restores identifying variation – the methodological logic formalized by Goldsmith-Pinkham et al. (2020) for Bartik-style instruments generally.

We instrument FireExposure_{it} with $\text{OniExposure}_{it} = w_i \times \text{ONI}_t$ in a two-stage least-squares specification, exploiting the plausibly exogenous influence of ENSO-driven dry conditions on Indonesian peatland fire activity (Hayasaka et al., 2020). Instrument strength is assessed via the first-stage F-statistic, with $F > 10$ taken as the conventional threshold for adequate strength.

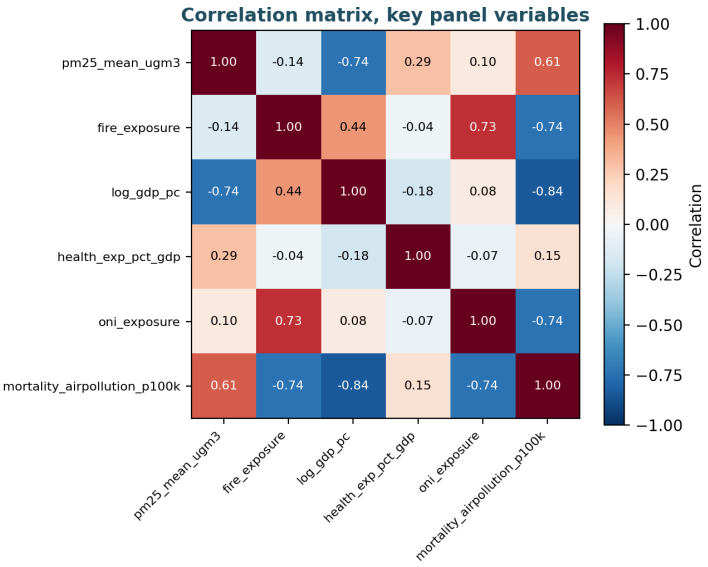


Figure 5: Correlation matrix, key panel variables.

3.4.2 Mortality (Cross-Sectional Estimation)

Because mortality data are non-missing only for 2019 (Section 3.2), Equation (7) cannot be estimated for this outcome: with one observation per entity, country fixed effects are not identified, and no within-country time variation exists for a lead/lag falsification test or for Driscoll-Kraay standard errors. We instead estimate a cross-sectional specification on the 2019 subsample,

$$\text{Mortality}_i = \beta_1 \cdot \text{FireExposure}_i + \beta_2 \cdot \log(\text{GDPpc})_i + \varepsilon_i \tag{8}$$

with heteroskedasticity-consistent HC3 standard errors, the recommended choice once n falls substantially below the conventional threshold of approximately 30 observations typically assumed for asymptotic cluster-robust inference (Cameron and Miller, 2015). We report three specifications (fire exposure alone; total PM2.5 alone; both jointly) and a leave-one-out jackknife, refitting the model with each country excluded in turn.

3.4.3 Inference with a Small Number of Clusters

A methodological issue specific to this panel's dimensions warrants explicit discussion. Conventional cluster-robust standard errors require a substantially larger number of clusters than the conventional practitioner floor of ten to be reliably valid (Cameron and Miller, 2015); our recipient subsample contains four countries and our domestic-channel subsample contains two. In an earlier specification, we applied standard cluster-robust inference to the two-cluster domestic subsample and obtained a standard error of essentially zero together with a spuriously significant coefficient – an artifact of the estimator's breakdown at this sample size rather than a substantive finding. We address this by applying heteroskedasticity-robust (non-clustered) standard errors to any specification with fewer than ten clusters, flagged explicitly in Table 5, following the diagnostic guidance in Cameron and Miller (2015) (Section VI).

4 Results

4.1 Theoretical Results

This section reports the numerical implications of Equations (3)–(5) under the baseline calibration $a = 10$, $b = 1$, $d = 4$, presented without interpretation; discussion is deferred to Section 5.3.

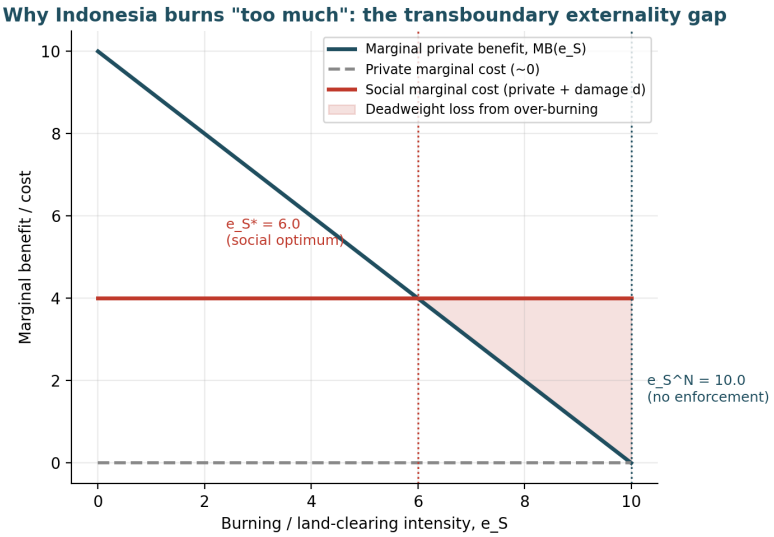


Figure 6: Marginal private benefit and marginal social cost of burning intensity, baseline calibration.

Table 2: Baseline calibration results

Quantity	Value
e_S^N (non-cooperative burning level)	10.0
e_S^* (socially optimal burning level)	6.0
Over-pollution gap ($e_S^N - e_S^*$)	4.0
π_S at e_S^N	50.0
π_S at e_S^*	42.0

Under the baseline calibration, Equations (3) and (4) yield $e_S^N = 10.0$ and $e_S^* = 6.0$, an over-pollution gap of 4.0 units (Table 2, Figure 6). Figure 7 and Table 3 report the comparative statics of this gap with respect to the damage parameter d.

Table 3: Comparative statics with respect to damage severity

Damage d	e_S^N	e_S^*	Gap	DWL (approx.)
0	10.0	10.0	0.0	0.0
2	10.0	8.0	2.0	2.0
4	10.0	6.0	4.0	8.0
6	10.0	4.0	6.0	18.0
8	10.0	2.0	8.0	32.0
10	10.0	0.0	10.0	50.0

As d increases from 0 to 10, e_S^N remains fixed at 10.0 by construction (Equation (3) does

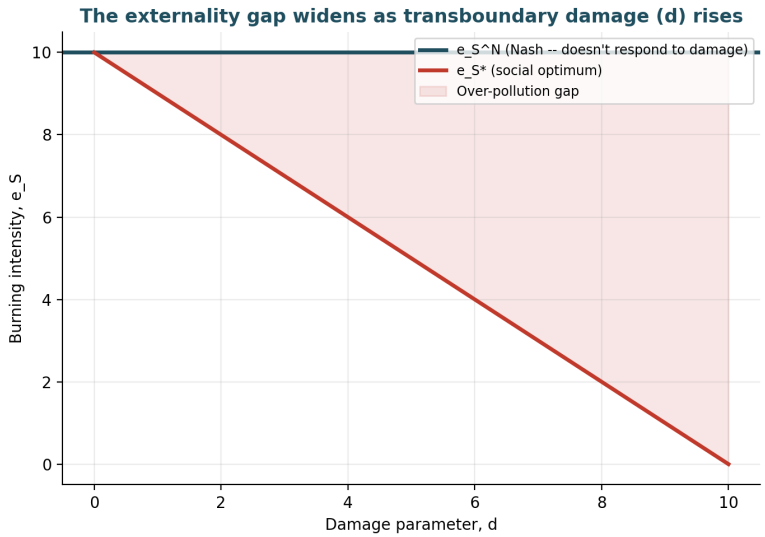


Figure 7: Comparative statics: e_S^* as a function of the damage parameter d.

not depend on d), while e_S^* declines linearly from 10.0 to 0.0, and the associated deadweight loss increases from 0.0 to 50.0 (Table 3).

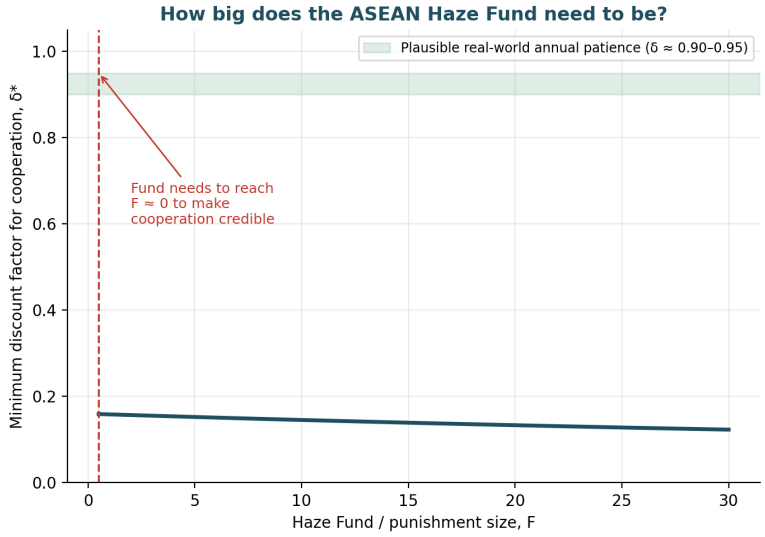


Figure 8: $\delta^*(F)$: minimum discount factor for cooperation as a function of fund size.

Table 4 and Figure 8 report $\delta^*(F)$ for $F \in \{1, 5, 10, 15, 20, 25, 30\}$. Under this calibration, $\delta^*(F)$ declines from 0.158 at $F = 1$ to 0.123 at $F = 30$, remaining below 0.92 – the reference value used throughout for a plausible annual discount factor – across the entire range considered.

Figure 9 depicts the (δ, F) combinations for which cooperation is sustainable under Equation (5); the plausible discount-factor range ($\delta \approx 0.90$–0.95) lies within the sustainable region across the full range of F considered. Figure 10 reports the three payoffs entering Equation (5) directly at $F = 10$: $\pi_S(e_S^N) = 50.0$, $\pi_S(e_S^*) = 42.0$, and $\pi_S^{\text{punish}}(10) = -5.0$.

Table 4: $\delta^*(F)$ across fund sizes

Fund size F	$\pi_S^{\text{punish}}(F)$	$\delta^*(F)$	Sustainable at $\delta = 0.92$?
1	-0.5	0.158	Yes
5	-2.5	0.152	Yes
10	-5.0	0.145	Yes
15	-7.5	0.139	Yes
20	-10.0	0.133	Yes
25	-12.5	0.128	Yes
30	-15.0	0.123	Yes

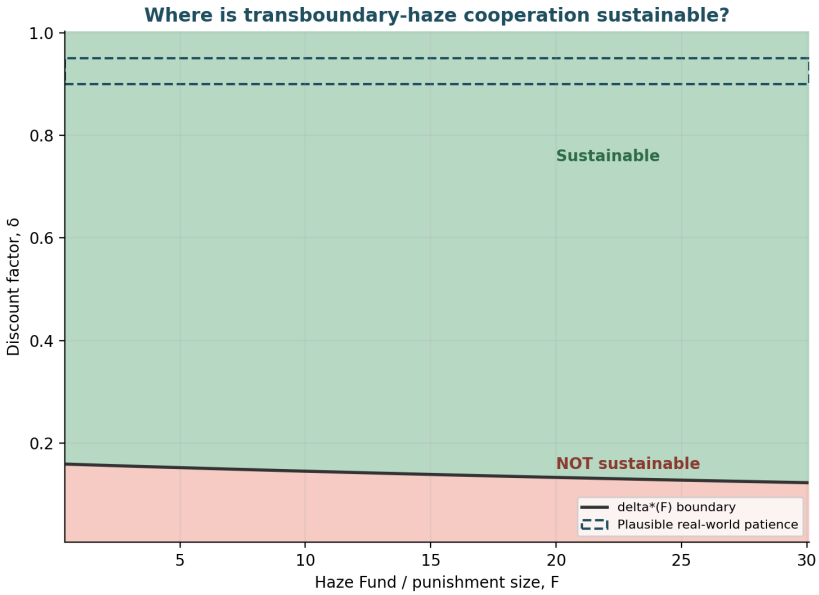


Figure 9: Feasibility region for cooperation in (δ, F) space.

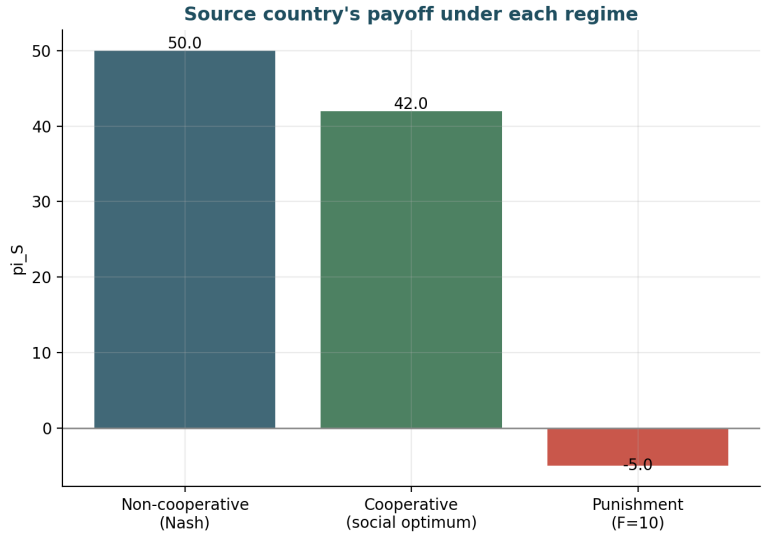


Figure 10: Source payoffs under the non-cooperative, cooperative, and punishment regimes ($F = 10$).



Figure 11: Indonesian fire activity and El Niño (ONI) intensity, 2015–2024.

4.2 Pollution Pathway Results

Figure 11 reports Indonesia's annual fire hotspot count alongside the fire-season ONI value for 2015–2024. The highest fire count in the sample (2015) coincides with the highest ONI value; the lowest fire counts (2020–2022) coincide with negative ONI values (La Niña conditions); a secondary fire peak in 2023 coincides with a secondary ONI peak.

Figure 12 reports mean PM2.5 by channel group over time. The recipient-group series exhibits visible peaks in the shaded El Niño years (2015, 2019); the domestic-channel series exhibits a comparatively monotonic upward trend not visibly associated with the shaded years.

Table 5: PM2.5 panel estimation results

Specification	Coefficient (fire_exposure)	N	Within R^2
(1) Baseline FE	0.0172*** † (0.0035)	36	0.432
(2) Excl. 2020–21	0.0162*** † (0.0038)	28	0.467
(3) IV / 2SLS	0.0167*** (0.0063)	36	n/a (2SLS)
(4) Domestic placebo	−0.0191 † (0.0164)	18	0.658

*/**/*** denote significance at the 10/5/1 percent level. † indicates heteroskedasticity-robust (non-clustered) standard errors, applied where the relevant subsample contains fewer than ten clusters (Section 3.4).

Table 5 and Figure 13 report four specifications. The baseline two-way fixed-effects estimate (column 1) is 0.0172 ($p < 0.01$). Excluding 2020–2021 (column 2) yields 0.0162 ($p < 0.01$). The instrumental-variables estimate (column 3) is 0.0167 ($p < 0.01$), with a first-stage F-statistic of 82.1 on the excluded instrument. The domestic-channel placebo (column 4) yields −0.0191, not statistically distinguishable from zero (standard error 0.0164).

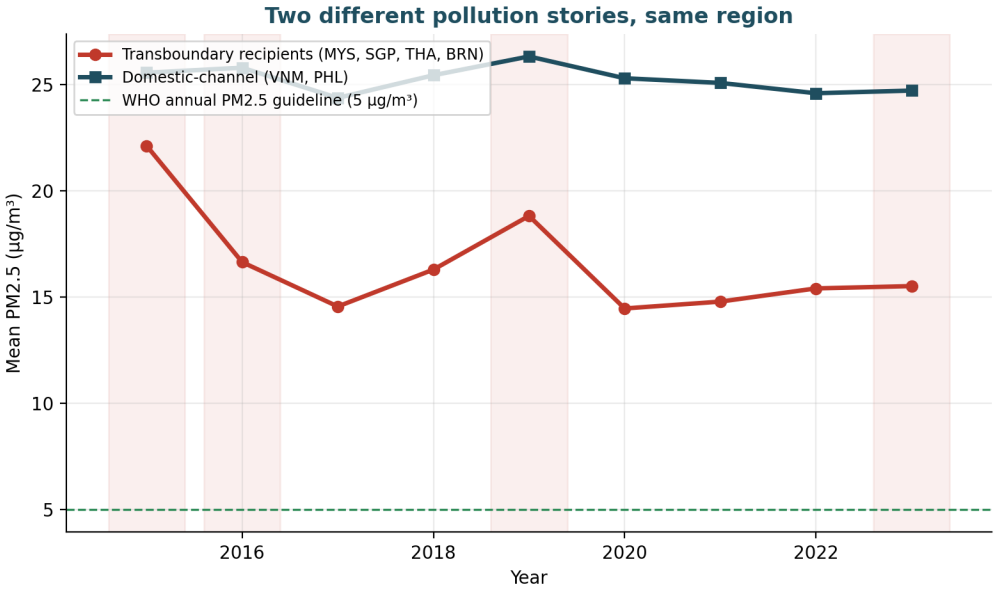


Figure 12: PM2.5 by channel group over time, El Niño years shaded.

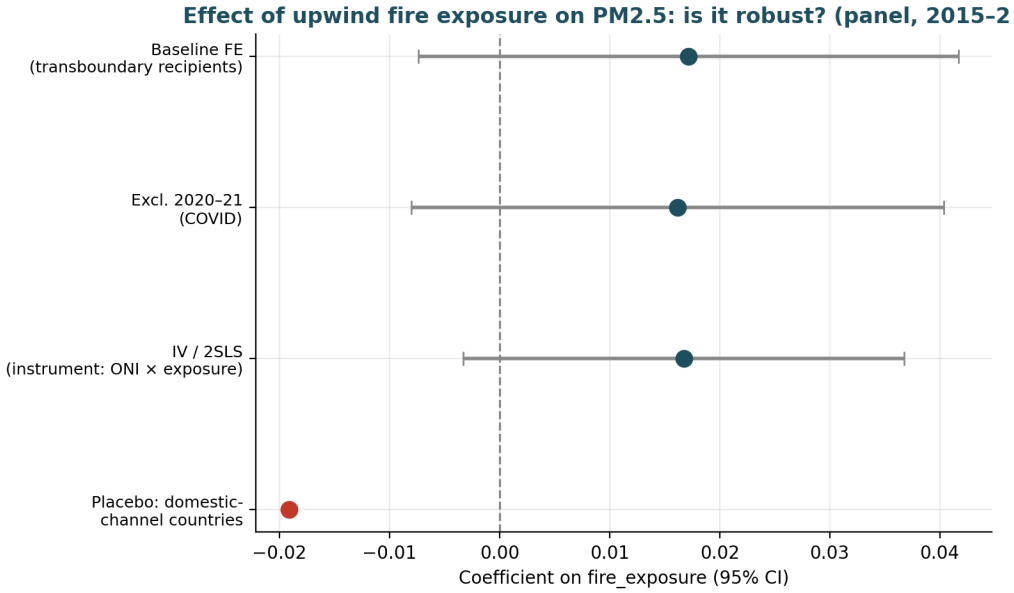


Figure 13: Estimated effect of fire exposure on PM2.5 across specifications.

Table 6: Mortality cross-section estimation results

Specification	Coefficient	N
(1) Fire exposure only	−0.0708 (0.1772)	10
(2) Total PM2.5 only	0.2972 (2.1548)	10
(3) Fire exposure, ctrl. PM2.5	−0.0757 (0.2122)	10
Leave-one-out range (spec 1)	[−0.2270, −0.0183]	9 (each)

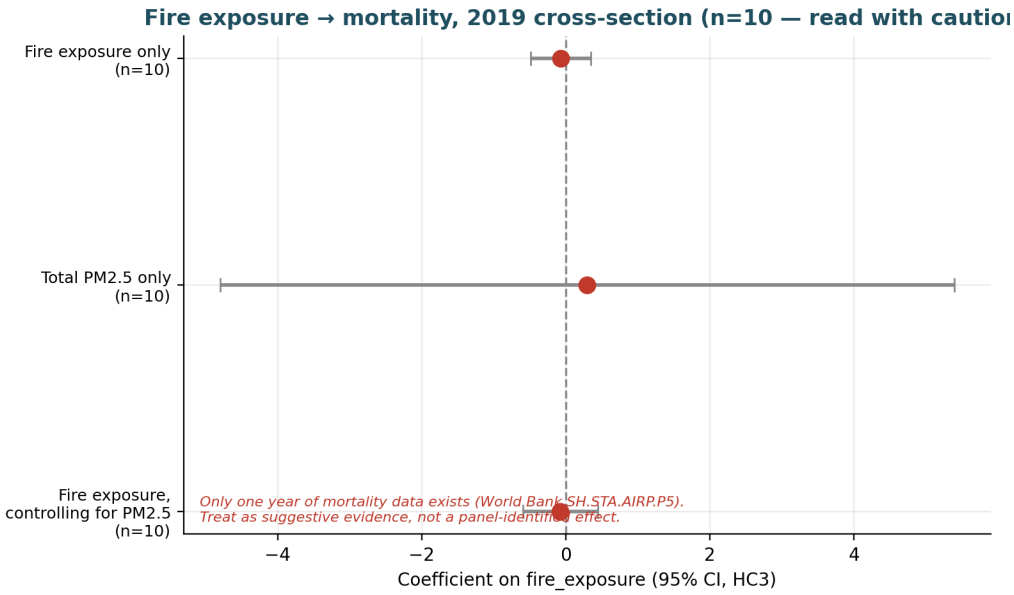


Figure 14: Estimated effect of fire exposure and PM2.5 on mortality, 2019 cross-section ($n = 10$).

4.3 Mortality Cross-Section Results

HC3 heteroskedasticity-robust standard errors throughout. None of the three specifications in Table 6 yields a coefficient statistically distinguishable from zero. The leave-one-out jackknife (final row, and Figure 15) produces estimates ranging from -0.2270 to -0.0183 across the ten country-exclusion refits, with the sign remaining negative throughout but no refit achieving statistical significance.

5 Discussion

5.1 The Pollution Pathway in Relation to Existing Estimates

The Table 5 result that fire exposure raises PM2.5 among transboundary-recipient countries is directionally consistent with, though smaller in magnitude than, the estimate reported by Sheldon and Sankaran (2017), who find that a one-standard-deviation increase in Indonesian fire radiative power raises Singapore's Pollution Standards Index by approximately 16 percent. We attribute this difference primarily to two design differences rather than to disagreement between the studies. First, Sheldon and Sankaran (2017) study Singapore alone, a small, geographically homogeneous, low-domestic-pollution setting in which the fire signal is comparatively unconfounded; our four-country recipient panel includes Thailand and Malaysia, both of which have non-trivial domestic pollution sources that plausibly attenuate the average estimated effect. Second, our fire measure is a simple detection count rather than the fire radiative power measure used by Sheldon and Sankaran (2017), which captures fire intensity rather than mere occurrence and would be expected to produce a more precisely estimated relationship. We interpret our contribution as complementary rather than competing: where Sheldon and Sankaran (2017) establish the relationship with high precision for a single, ideal case, we establish that the same qualitative relationship extends, with reduced but still statistically detectable magnitude, across

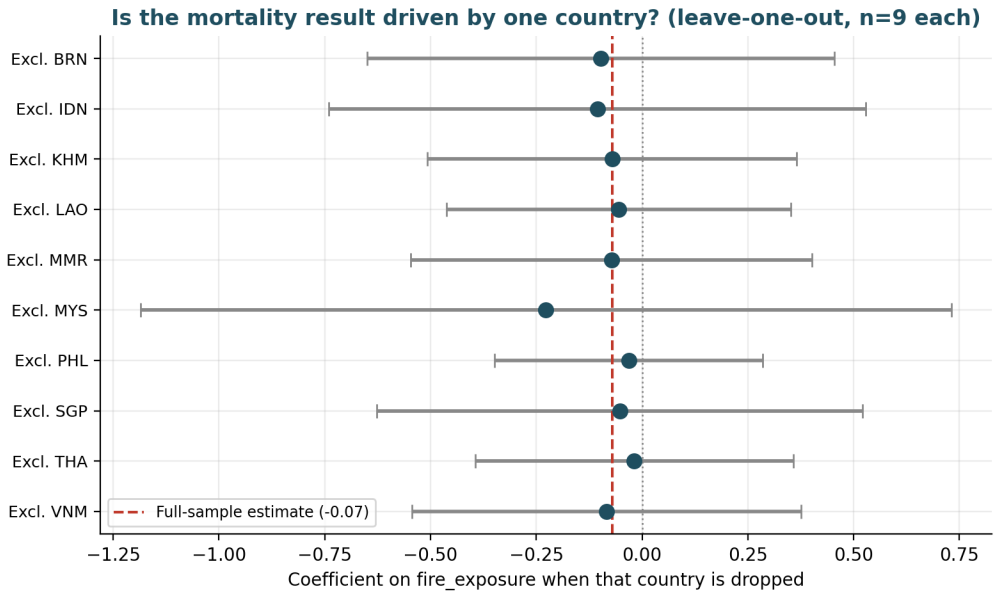


Figure 15: Leave-one-out jackknife: coefficient on fire exposure when each country is excluded in turn.

the full set of transboundary-recipient states, and we connect this empirical relationship to the compensation-fund model in Section 3.1, which Sheldon and Sankaran (2017) do not address.

5.2 The Mortality Null Result in Context

The absence of a statistically significant mortality effect in Table 6 should not be interpreted as evidence against the substantial health effects documented elsewhere in the literature. Koplitz et al. (2016) estimate approximately 100,300 excess deaths attributable to the 2015 haze episode alone using chemical transport modeling with substantially richer spatial and temporal resolution than the single annual cross-section available to us; Jayachandran (2009) similarly identifies a large and statistically robust mortality effect using a research design (missing-cohort analysis applied to census microdata) that does not depend on the year-to-year variation our cross-sectional design requires. Our null result is, in our assessment, a direct consequence of statistical power – a ten-observation cross-section with approximately six to seven degrees of freedom, depending on specification, is not capable of detecting an effect of the magnitude these studies report, even if that effect is present in our data. We regard Table 5's well-identified pollution-pathway result as evidence that the first link in the causal chain these mortality studies presume (fires raise pollution) is empirically supported in our broader ten-country setting; we were simply unable to extend that chain to mortality with the single year of data available.

5.3 The Cooperation Puzzle: Theoretical Prediction versus Institutional Reality

Section 4.1 established that, under our baseline calibration, $\delta^*(F)$ remains below 0.16 across the full range of fund sizes considered – substantially below the plausible discount-factor range of 0.90–0.95 – implying that cooperation should be readily sustainable even at modest fund levels. This prediction is difficult to reconcile with the AATHP's documented history of chronic

underfunding and recurring non-compliance (Section 2.3). We do not consider it appropriate to recalibrate the model until this discrepancy disappears, as doing so would convert a genuine empirical puzzle into an artifact of parameter choice. Instead, we interpret the discrepancy as informative about the model's central simplifying assumption: the treatment of the source as a single, unitary decision-maker. The institutional literature reviewed in Section 2.3 (Muhammad, 2022; Chaves et al., 2025; Ryu and Lee, 2026; Pham and Thomson, 2025) attributes the AATHP's weakness to ASEAN's non-interference norm and its members' broader reluctance to delegate discretionary power, which undermines the credibility of the grim-trigger punishment threat central to Equation (5) – formally, this corresponds to p remaining close to zero regardless of nominal fund size, since the punishment is never credibly executed. A complementary explanation, external to the institutional literature but consistent with it, is that the burning decision is in practice made by a large number of independent smallholders and plantation operators rather than a single national actor, so that a national-level compensation mechanism provides negligible incentive to any individual decision-maker – a collective-action problem structurally distinct from the two-player game in Section 3.1. We regard disaggregating the source into multiple heterogeneous agents as the most promising extension of the present model and do not attempt it here.

5.4 Methodological Considerations for Small Regional Panels

The small-cluster inference issue documented in Section 3.4 is not specific to the domestic-channel placebo in which it was first identified; it applies in principle to the four-country recipient subsample as well, and Table 5 applies the same correction throughout. We highlight this methodological point because ten-member regional panels of the kind studied here are common in Southeast Asian and broader regional economic research, and the failure mode we encountered – a superficially precise but statistically invalid standard error – is, in our reading of Cameron and Miller (2015), a general risk in this class of study rather than one specific to our design. We recommend that researchers working with regional panels of comparable size report cluster counts explicitly and apply the diagnostic checks summarized in Cameron and Miller (2015) (Section VI) before reporting cluster-robust inference.

5.5 Limitations

- The exposure weights described in Section 3.2 are illustrative rather than estimated; a more defensible construction – for example, based on historical correlation between each country's ground-station PM2.5 readings and Sumatran fire activity, or wind back-trajectory frequency – would strengthen the quantitative interpretation of Table 5's coefficients.
- The panel is restricted to annual frequency and to 2015–2023 for PM2.5 (2024 not yet published by the World Bank at the time of writing); a monthly or fire-season-restricted panel would increase statistical power without requiring additional years of data.
- Thailand is classified entirely as a transboundary recipient, though Northern Thailand's own agricultural burning also contributes to its measured PM2.5 (Section 2.2); this simplification most likely biases the estimated recipient-group effect toward zero for this specific country rather than inflating it.

- MODIS is used only as a fallback for a 2022 coverage gap in the VIIRS-SNPP series; a formal sensitivity analysis re-estimating Table 5 with MODIS as the primary fire series is a natural robustness extension not undertaken here.
- The mortality analysis (Section 4.3) is, as discussed in Section 5.2, underpowered by design given available data; we regard it as suggestive rather than conclusive.

6 Conclusion and Policy Recommendations

This paper developed a repeated-game model of transboundary haze pollution, calibrated to the documented funding history of the ASEAN Transboundary Haze Pollution Control Fund, and tested its core empirical prediction using a novel ten-country, ten-year panel constructed from satellite fire data and World Bank economic and environmental indicators. We find robust evidence that fire exposure raises PM2.5 concentrations among transboundary-recipient countries, a result that survives instrumental-variable estimation and a domestic-pollution placebo test. We do not find a statistically significant mortality effect, a limitation of available data rather than evidence against the underlying health relationship documented elsewhere in the literature. We further document a theoretical tension between the model's prediction that cooperation should be readily sustainable and the empirical persistence of non-compliance, which we interpret as evidence that collective-action and enforcement-credibility considerations, external to the model's single-agent representation of the source country, are the binding constraint on regional cooperation.

Subject to the limitations discussed in Section 5.5, we draw three policy implications from these findings.

1. Fund size alone is unlikely to be the binding constraint on cooperation. Section 4.1's results imply that cooperation should already be sustainable at fund levels well below the AATHP's actual (still inadequate) funding; policy attention directed solely at fund size, without addressing enforcement credibility, is therefore unlikely to close the compliance gap identified in Section 5.3.
2. Because the burning decision is made in practice by numerous independent land-clearing actors rather than a single national authority (Section 5.3), satellite-based fire attribution at the level of individual land concessions – technically feasible using the same FIRMS data employed in this study – represents a more direct policy lever than aggregate, country-level compensation.
3. External supply-chain regulation, such as the European Union's deforestation-linked import restrictions, may substitute effectively for weak intra-ASEAN enforcement mechanisms by altering the incentives of individual plantation operators directly, circumventing the non-interference constraint identified in the institutional literature (Section 2.3).

A Formal Derivations

A.1 Deadweight Loss from Over-Burning

Given $e_S^N = a/b$ and $e_S^* = (a-d)/b$ from Equations (3)–(4), the deadweight loss from over-burning is the area between the marginal-benefit curve $MB(e_S) = a - b\,e_S$ and the marginal social-cost line $MC^{\text{social}} = d$, evaluated between e_S^* and e_S^N:

$$DWL = \int_{e_S^*}^{e_S^N} [MB(e_S) - d]\, de_S = \int_{e_S^*}^{e_S^N} [a - b\,e_S - d]\, de_S \tag{9}$$

Since $a - b\,e_S - d = 0$ at $e_S = e_S^*$ by construction (Equation (4)), the integrand is linear and vanishes at the lower limit, so the integral reduces to the area of a right triangle with base $(e_S^N - e_S^*) = d/b$ and height $b(e_S^N - e_S^*) = d$, giving:

$$DWL = \frac{1}{2}(e_S^N - e_S^*) \cdot d = \frac{1}{2}\left(\frac{d}{b}\right) d = \frac{d^2}{2b} \tag{10}$$

At the baseline calibration ($b = 1$, $d = 4$), $DWL = 16/2 = 8$, matching the value reported in Table 3 for $d = 4$. Table 3's remaining rows follow by direct substitution.

A.2 Derivation of the Cooperation Condition (Equation (5))

The value of perpetual cooperation, starting in the current period, is the discounted sum of the stage payoff $\pi_S(e_S^*)$ received in every period:

$$V^C = \pi_S(e_S^*) + \delta\,\pi_S(e_S^*) + \delta^2\,\pi_S(e_S^*) + \cdots = \pi_S(e_S^*) \sum_{i=0}^{\infty} \delta^i = \frac{\pi_S(e_S^*)}{1-\delta} \tag{11}$$

The value of deviating in the current period – playing the myopic best response e_S^N once, then receiving the fixed punishment payoff π_S^{punish} in every subsequent period under the grim-trigger strategy – is:

$$V^D = \pi_S(e_S^N) + \delta\,\pi_S^{\text{punish}} + \delta^2\,\pi_S^{\text{punish}} + \cdots = \pi_S(e_S^N) + \frac{\delta\,\pi_S^{\text{punish}}}{1-\delta} \tag{12}$$

Cooperation is sustainable in a subgame-perfect equilibrium if and only if $V^C \geq V^D$:

$$\frac{\pi_S(e_S^*)}{1-\delta} \geq \pi_S(e_S^N) + \frac{\delta\,\pi_S^{\text{punish}}}{1-\delta} \tag{13}$$

Multiplying both sides by $(1-\delta) > 0$:

$$\pi_S(e_S^*) \geq (1-\delta)\pi_S(e_S^N) + \delta\,\pi_S^{\text{punish}} = \pi_S(e_S^N) - \delta\,\pi_S(e_S^N) + \delta\,\pi_S^{\text{punish}} \tag{14}$$

Rearranging to isolate δ:

$$\delta\left[\pi_S(e_S^N) - \pi_S^{\text{punish}}\right] \geq \pi_S(e_S^N) - \pi_S(e_S^*) \tag{15}$$

Since $\pi_S(e_S^N)$ is the unconstrained maximum of $\pi_S(\cdot)$ and π_S^{punish} is assumed strictly worse (a maintained assumption verified numerically in Section 4.1), the bracketed term is strictly positive, permitting division without a sign change:

$$\delta \geq \frac{\pi_S(e_S^N) - \pi_S(e_S^*)}{\pi_S(e_S^N) - \pi_S^{\text{punish}}} \equiv \delta^*(F) \tag{16}$$

which is Equation (5) in the main text. Substituting $\pi_S^{\text{punish}}(F) = -\kappa F$ and the closed-form payoffs $\pi_S(e_S^N) = a^2/(2b)$ and $\pi_S(e_S^*) = (a-d)(a+d)/(2b)$ [obtained by substituting Equations (3)–(4) into the payoff function $a\,e_S - (b/2)e_S^2$] yields $\delta^*(F)$ explicitly as a function of the primitive parameters (a, b, d, κ, F), which we evaluate numerically in Table 4 rather than report in closed form given the number of parameters involved.

A.3 Comparative Statics of $\delta^*(F)$ with Respect to Damage Severity

Differentiating $\delta^*(F)$ with respect to d (holding a, b, F, κ fixed) shows that the numerator $\pi_S(e_S^N) - \pi_S(e_S^*)$ is increasing in d (since $\pi_S(e_S^*)$ is decreasing in d, as e_S^* itself is decreasing in d), while the denominator $\pi_S(e_S^N) - \pi_S^{\text{punish}}(F)$ does not depend on d. It follows immediately that $\delta^*(F)$ is increasing in d: holding fund size fixed, cooperation becomes strictly harder to sustain as transboundary damage becomes more severe, since the cooperative payoff $\pi_S(e_S^*)$ that the source forgoes by not deviating falls further below the deviation payoff $\pi_S(e_S^N)$. This comparative static is not separately tabulated in the main text but follows directly from Equation (5) and Table 3.

References

Barrett, S. (2001). Strategic environmental policy and international trade. In *The Economics of International Trade and the Environment* (pp. 83–93). CRC Press.

Baumol, W. J., & Oates, W. E. (1988). *The theory of environmental policy*. Cambridge University Press.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. *Journal of Human Resources*, *50*(2), 317–372.

Castro, M., Ritonga, R., Masuda, Y., Asyhari, A., Novita, N., Gangga, A., … & Leavitt, S. (2026). Effective restoration can avoid peatland fires: Large scale counterfactual assessment in Kalimantan, Indonesia. *iScience*, *29*(6).

Chaves, I. V., Quessep-Feria, J. R., Narváez-Mercado, B., & Valdelamar-Montes, J. A. (2025). A supranational law perspective on the right to a healthy environment in Southeast Asia: The ASEAN agreement on transboundary haze pollution. *Fronteira: Journal of Social, Technological and Environmental Science*, *14*(4), 194–207.

Coase, R. H. (2013). The problem of social cost. *The Journal of Law and Economics*, *56*(4), 837–877.

Goldsmith-Pinkham, P., Sorkin, I., & Swift, H. (2020). Bartik instruments: What, when, why, and how. *American Economic Review*, *110*(8), 2586–2624.

Hayasaka, H., Usup, A., & Naito, D. (2020). New approach evaluating peatland fires in Indonesian factors. *Remote Sensing*, *12*(12), 2055.

Jayachandran, S. (2009). Air quality and early-life mortality: Evidence from Indonesia's wildfires. *Journal of Human Resources*, *44*(4), 916–954.

Kiely, L., Spracklen, D. V., Wiedinmyer, C., Conibear, L., Reddington, C. L., Arnold, S. R., et al. (2020). Air quality and health impacts of vegetation and peat fires in equatorial Asia during 2004–2015. *Environmental Research Letters*, *15*, 094054.

Koplitz, S. N., Mickley, L. J., Marlier, M. E., Buonocore, J. J., Kim, P. S., Liu, T., Sulprizio, M. P., DeFries, R. S., Jacob, D. J., Schwartz, J., Pongsiri, M., & Myers, S. S. (2016). Public health impacts of the severe haze in Equatorial Asia in September–October 2015: Demonstration of a new framework for informing fire management strategies to reduce downwind smoke exposure. *Environmental Research Letters*, *11*(9), 094023.

Muhammad, F. (2022). Environmental agreement under the non-interference principle: The case of ASEAN agreement on transboundary haze pollution. *International Environmental Agreements: Politics, Law and Economics*, *22*(1), 139–155.

Pham, H., & Thomson, R. (2025). The delegation of discretionary power in international agreements: New comparative evidence from the Association of Southeast Asian Nations. *International Relations of the Asia-Pacific*, *25*(1), 1–37.

Pigou, A. (2017). *The economics of welfare*. Routledge.

Putra, B. A. (2024). The politics of environmental policy: Haze pollution, ASEAN, and the way forward. *Frontiers in Sustainable Cities*, *6*, 1417746.

Ryu, S. H., & Lee, J.-W. (2026). ASEAN's adaptive institutional responses to transboundary haze pollution. *SUVANNABHUMI*, *18*(1), 135–170.

Sastry, N. (2002). Forest fires, air pollution, and mortality in Southeast Asia. *Demography*, *39*(1), 1–23.

Sheldon, T. L., & Sankaran, C. (2017). The impact of Indonesian forest fires on Singaporean pollution and health. *American Economic Review*, *107*(5), 526–529.

Sheldon, T. L., & Sankaran, C. (2019). Averting behavior among Singaporeans during Indonesian forest fires. *Environmental and Resource Economics*, *74*(1), 159–180.

World Bank. (2016). *The cost of fire: An economic analysis of Indonesia's 2015 fire crisis*. World Bank.

Yu, C., Tu, G., & Yu, F. (2024). Research on transboundary air pollution control and cooperative strategies based on differential game. *Atmosphere*, *15*(12), 1528.