Working paper · Environmental economics · June 2026

International Environmental Agreements Agreements with Minimum Participation Participation Thresholds

How a binding minimum abatement, combined with a minimum participation rule, can reconcile broad inclusiveness with deep ambition in global climate cooperation.

Presented at CRETE 2026
The 24th Conference on Research on Economic Theory and Econometrics
Ios, July 07-11, 2026

George Deltas *University of Illinois*

Eftichios Sartzetakis *University of Macedonia*

Climate cooperation is a global public-good problem

Benefits of abatement are global; costs are local. Free-riding drives persistent under-provision.

The puzzle

Weak commitments, limited participation

Most IEAs deliver only marginal welfare gains above the Nash outcome — Montreal is the exception, not the rule.

The instrument

Participation thresholds

A treaty enters into force only when a critical mass ratifies — a coordination device against the futility concern.

Real-world use

Kyoto & Paris both use them

Kyoto: 55% of Annex I emissions. Paris: 55 countries & 55% of global emissions.

Why study them now

Thresholds are *institutionally central* yet *theoretically under-explored*: we still lack a unified account of how they jointly shape ambition and participation.

Motivating question: What does a threshold treaty *really* buy us?

> *How do **minimum participation thresholds** — combined with a binding minimum abatement target— alter equilibrium coalition size, total abatement, and global welfare?*

Four contributions

i. Unified framework

Tractable model with heterogeneous valuations, a common cost, an abatement floor, *and* a participation threshold.

ii. Non-monotonic tradeoff

Ambition and participation co-rise at first, then trade off — the feasible frontier is strictly concave.

iii. Median-voter result

A sequential median-voter process replicates the constrained welfare optimum — democracy aligns with efficiency.

iv. Quantitative calibration

Using LSCC (Cruz & Rossi-Hansberg, 2024) with population and emissions data — coverage >90% of global emissions.

Headline — threshold treaties can deliver 80–90% of first-best welfare gains.

Three strands — one synthesis

I Self-enforcement

Self-enforcing IEAs are small

Hoel (1992); Carraro & Siniscalco (1993).

Barrett (1994): self-enforcement severely caps coalition size.

Diamantoudi et al. (2026): meaningful cooperation needs cost-effective technology.

Conclusion — coalitions deliver only marginal welfare gains above Nash.

II Stabilizing devices

Mechanisms that expand cooperation

Chander & Tulkens (1995, 1997) — transfers and cost-sharing.

Farsightedness (de Zeeuw 2008; Diamantoudi & Sartzetakis 2015).

Club goods: technology (Barrett 2006), trade access (Eichner & Pethig).

Repeated games (Dutta & Radner 2004; Harstad 2020).

Conclusion — narrow-but-deep deals work when enforcement is real.

III Threshold rules · *our home*

Participation & threshold design

Finus & Rundshagen (2006); Asheim et al. (2006).

McGinty (2007, 2011); Rubio & Ulph (2007).

Finus & Maus (2008) — thresholds stabilise coalitions.

Weikard & Dellink (2014) — support more ambitious targets.

Carraro, Marchiori and Oreffice (2009) ERE
Weikard, Wangler and Freytag (2015) ERE
Köke and Lange (2017) JEEM

Free-riding problem: if a country stays out of an agreement
- it still gets most of the benefits, due to the actions of all other members.
- but does not have to pay any of the costs.

We add joint analytics of ambition & participation, plus quantitative calibration.

The Model: Heterogeneous valuations, shared cost

Setup

n countries, indexed $i = 1, \ldots, n$. Each chooses abatement $a_i \geq 0$.

Total abatement $A = \sum_i a_i$.

Linear global benefit, quadratic local cost — the workhorse public-good payoff.

Country welfare:

$$w_i = \beta_i A - \tfrac{1}{2} c a_i^2$$

linear benefit · quadratic cost · common cost parameter c

Heterogeneity

Countries differ only in β_i.
Rank so $\beta_1 > \beta_2 > \ldots > \beta_n$, with constant gap $\Delta\beta$ so that $\beta_1 = \beta_{max}$.
Constraint: $\beta_{max} / \Delta\beta > n - 1$ (all benefits positive).

Four-stage game

i. **Choose a^s**
Minimum abatement required of all signatories.

ii. **Choose k^s**
Minimum participation threshold for entry-into-force.

iii. **Sign or not**
Each country decides whether to ratify, given (a^s, k^s).

iv. **Abate**
Signatories abate ≥ a^s if threshold met; others optimise privately.

Solution concept

Backward induction. Focus on the equilibrium where the k highest-β countries sign.

Two levers: a minimum abatement and a minimum participation

Lever A · Ambition

a^s

Minimum abatement required of all signatories.

Parametrized as $a^\mu = \mu\,\beta_{max}/c$, with $\mu \geq 1$.
$\mu = 1 \Rightarrow$ minimal binding agreement.
Higher $\mu \Rightarrow$ deeper cuts, fewer signatories.

Lever B · Breadth

k^s

Minimum number of countries that must ratify.

Treaty fails — everyone reverts to Nash — if fewer than k^s ratify. The pivotal country *knows* it is pivotal.

How the threshold works

i. Anchors mutual expectation

No one ratifies alone — ratification only matters if others ratify too.

ii. Removes the futility excuse

Each pivotal country knows its vote tips the treaty in — not into the void.

iii. Expands the feasible set

Deeper cuts become incentive-compatible than under pure voluntarism.

The two reference points

No agreement

Nash equilibrium

Each country maximises its own welfare, taking others as given. The first-order condition gives:

$$a_i^N = \beta_i / c$$

Each country internalises only its own marginal benefit.
Total Nash abatement
$A^N = \frac{n}{2c}[2\beta_{max} - (n-1)\Delta\beta] \ll$ first-best.

Grand coalition

First best

Social planner maximises aggregate welfare; common solution because c is shared:

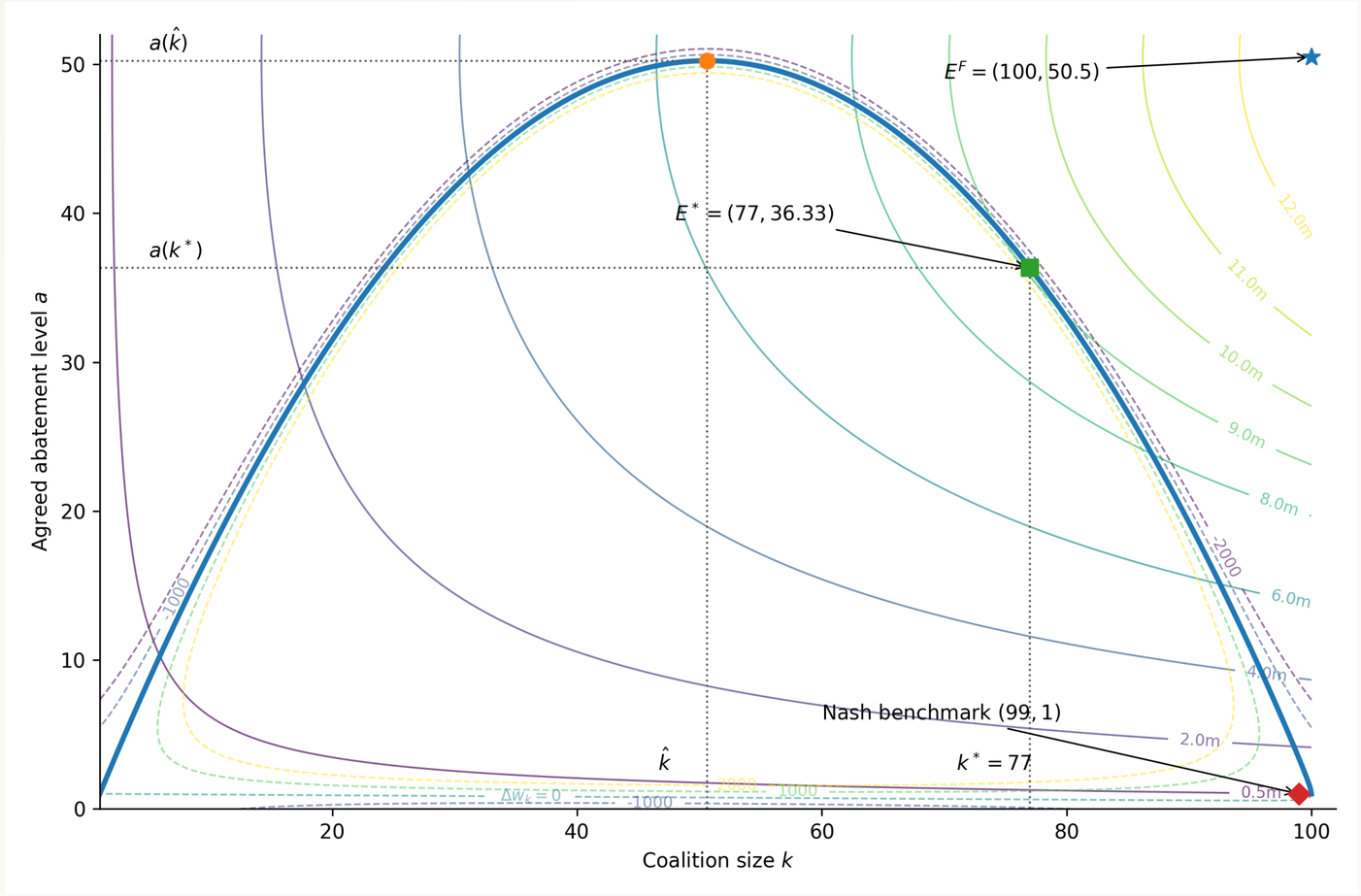
$$a_i^F = \left(\sum_{j=1}^{n} \beta_j\right) / c$$

Each country abates at the same rate; aggregate $A^F = n \cdot A^N$ — an n-fold gap.

The gap to close

Self-enforcement collapses the achievable region toward Nash. The treaty's job is to walk us up the line from A^N to A^F.

A non-monotonic, strictly concave frontier



Proposition 1

The frontier $a^{\mu}(k)$ is strictly concave on $(k-, k+)$.

Below $\hat{k}$. Tighter participation pulls in *more* ambition *and* more participants.

Above $\hat{k}$. Further tightening deters the pivotal country — ambition must fall.

Implication. The welfare-maximising treaty lives on the *downward-sloping* branch — small enough to be deep, big enough to matter.

Solve $w_k^{\mu} = w_k^N$ for a^{μ} which yields: $a^{\mu} = \Delta\beta \frac{k\Phi + \sqrt{(k-1)\Phi[(k-1)\Phi - k]}}{c}$, where $\Phi = \frac{\beta_{max}}{\Delta\beta} - (k-1)$

Bounds: $k\pm = \frac{B+1 \pm \sqrt{(B+1)(B-3)}}{2}$ *with* $B \equiv \beta_{max}/\Delta\beta$.

Smaller, deeper deals beat broader, shallower ones

Numerical illustration — $n = 100$, $\beta_{max} = 100$, $\Delta\beta = 1$, $c = 100$.

Minimal binding

$(a^s, k) = (1, 99)$

~ 0.5 Million

global welfare ≈ 495,050

99 countries sign at the Nash-of-the-largest-country abatement level. Lots of participation, almost no ambition.

Constrained optimum E*

$(a^*, k^*) \approx (36, 77)$

~ 9.06 M

~18× the minimal-binding deal

77 signatories abate $\mu^* \beta_{max}/c$ — the constrained first-best on the frontier.

Unconstrained first-best

$(a^F, k) = (50.5, 100)$

~ 12.8 M

SW = 12,751,250 — planner's ceiling

(1, 100) is infeasible — country 100 strictly prefers Nash. A threshold rule alone cannot deliver first-best.

Insight

The constrained optimum sits where the frontier's downward branch is tangent to the welfare contour — a far better deal than near-universal but minimal cooperation.

Democracy replicates the constrained optimum

Sequential voting · median rule

Stage i · pick a^v

Vote on the abatement level

Each country anticipates the threshold that will follow.

Stage ii · pick k^v

Vote on the threshold

Given a^v, all countries vote on minimum participation.

Stage iii · ratify

Sign without commitment

Each country joins iff joining beats Nash, given (a^v, k^v).

Key results

Lemma 1

Country preferences pile up at a single mass point: k_l — the largest threshold compatible with a coalition at a^v. All other preferred values are isolated singletons.

Propositions 3&4

The median voter is never pivotal — so the voting outcome is always $k^v = k_l$. The frontier is hit, not undershot. And in the first stage, the final agreement is $\{a^v(\theta), k_l(a^v(\theta);\theta)\}$

Main result

The median country's sequential vote on (a, k) almost reproduces the constrained welfare optimum — the same (a^*, k^*) a benevolent planner would pick on the frontier.

Voting results demonstrated

Stage ii · pick k^v

Let the vector of primitive parameters be $\theta \equiv (\beta_{max}, \Delta\beta, c)$, all assumed positive.
Consider a coalition of size k requiring each signatory to abate a
Define: w_i^N: welfare at Nash; $w_i^I(a,k;\theta)$ welfare of a coalition member, $w_i^O(a,k;\theta)$ welfare of a free-rider

Then:
A coalition of size k is sustainable at a^v if and only if $k \in [k_h(a^v;\theta), k_l(a^v;\theta)]$, where $k_h(a^v;\theta)$ and $k_l(a^v;\theta)]$ are, respectively, the smallest and largest coalition sizes that satisfy the participation constraint of the marginal member.

Define: $P(a^v;\theta)$ the # of countries that prefer a threshold below $k_l(a^v;\theta)$ so that they free ride on a coalition of size equal to their rank minus one. Also define $k_c(a^v;\theta)$ as the last country inside the sustainable coalition. Thus: $P(a^v;\theta) = k_l(a^v;\theta) - k_c(a^v;\theta)$.

Proposition 3.
For any abatement level a^v and admissible parameter vector θ:

$$P(a^v;\theta) \leq n/2$$

That is, in the linear-quadratic specification with linearly declining benefits, the number of countries that will not vote for a^v is weakly bounded by one half of the electorate.

Voting results demonstrated

Stage i · pick a^v

In the first stage countries do not vote over arbitrary pairs *(a,k)*, they rather vote over abatement proposals while anticipating that each proposal induces the agreement $\{a^v, k_l(a^v;\theta)\}$

Define: $$V_i(a;\theta) = \begin{cases} w_i^I(a, k_l(a;\theta);\theta), & \text{if } i \le k_l(a;\theta), \\ w_i^O(a, k_l(a;\theta);\theta), & \text{if } i > k_l(a;\theta). \end{cases}$$

the induced first-stage payoff of country i from proposal *a.*
Define also, $a_i^*(\theta) \in argmax V_i(a;\theta)$ as country *i*'s induced ideal abatement level

Then we prove:

Proposition 4
Suppose that countries vote sincerely over induced ideal abatement levels in the first stage. Then the sequential median-voter procedure selects

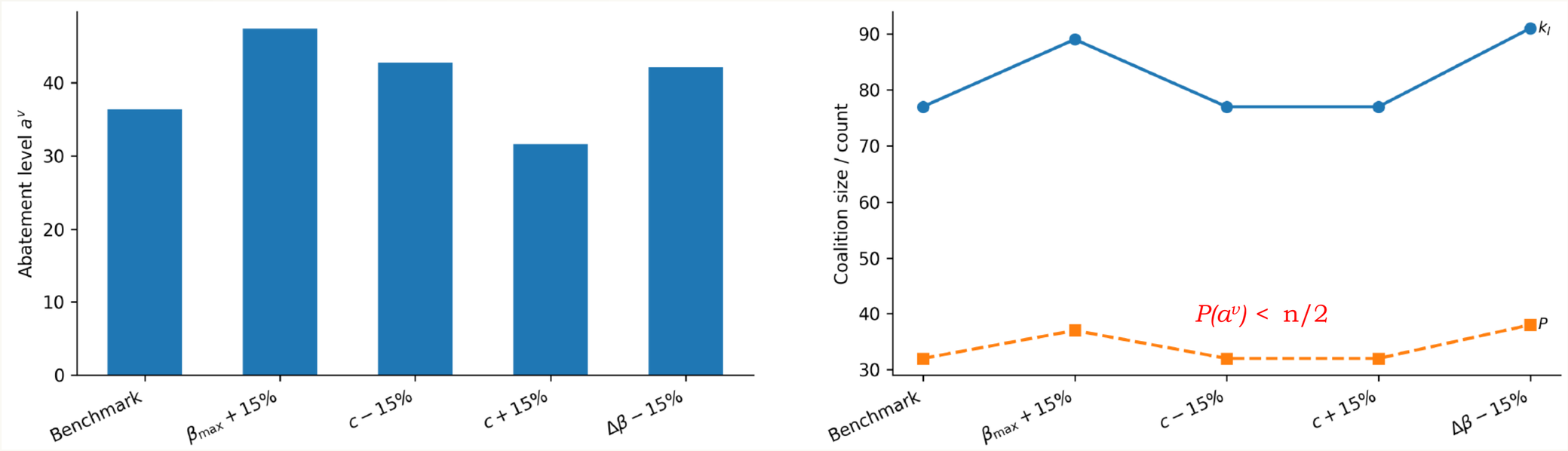
$$a^v(\theta) \in med\{a_1^*(\theta), \dots, a_n^*(\theta)\}, \qquad k^v(\theta) = \mathrm{k_l(av;\theta)}$$

Thus, the final agreement is

$$\{a^v(\theta), \mathrm{k_l}(a^v(\theta);\theta)\}.$$

Voting results: sensitivity

For each parameter vector □, we compute k, and then $a^v(k; □)$. Note that k_l changes as well.

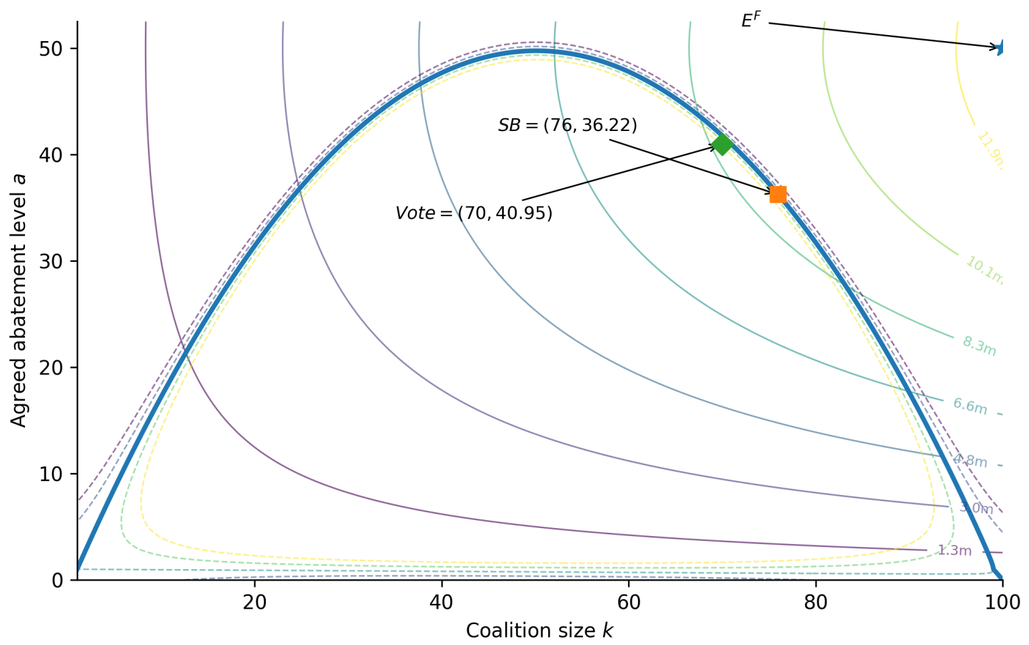


The qualitative median-voter result is robust: $P(a^v) < n/2$ and thus, $k^v = k_l$.

1. Increasing β_{max} by 15% raises both k^v and $a^v(k^v)$
2. Reducing $\Delta\beta$ by 15%, which makes countries more homogeneous, has a similar effect
3. Changing c rescales the sustainable abatement level without changing the endogenous coalition size.

Voting results: robustness and scope

Robustness



Constrained second-best and voting equilibrium under the highest admissible benefit heterogeneity (highest $\Delta b/b_{max}$) in the linear model

The voting mechanism remains quantitatively close to the second-best even under extreme admissible heterogeneity.

Although the distance between the two outcomes is larger than in the benchmark, the voting outcome still achieves approximately 97.6% of constrained second-best welfare.

The linear-quadratic model creates a smooth and single-peaked participation frontier, and the second-stage voting result implies that each first-stage abatement proposal is paired with the largest sustainable coalition.

Scope of the voting result

(1) Clustered benefits: a large group of low-benefit countries free riding on high-benefit countries, breaks $P(a) < n/2$.
(2) Different abatement costs: the ranking by benefits alone no longer identifies the marginal participant.
(3) Allowing for differentiated obligations, transfers, exemptions, or richer burden-sharing rules, the policy space becomes multidimensional and the direct median rule over a single abatement level is no longer sufficient.
(4) Strategic voting or agenda control could also move the outcome away from the sincere median of induced ideal points.

From toy model to real countries

Data sources (2022)

i. **Local Social Cost of Carbon (LSCC)**
Cruz & Rossi-Hansberg (2024a, 2024b). Mean $37 (Baseline) vs. $71 (High Damage). Cross-scenario correlation 0.98 — but rank shifts matter.

ii. **Population & CO_2 emissions per country**
World Development Indicators. Used to map LSCC $\sigma_i \mapsto$ per-capita valuation $\beta_i = \sigma_i e_i / (p_i E)$.

iii. **Abatement-cost parameter $c = 400$**
Calibrated from two technologies: direct-air carbon capture (~$400/ton at $\alpha = 1$) and solar-plus-storage substitution (~$80/ton at $\alpha = 0.5$). Both imply $c \approx 400$.

Model adjustments for heterogeneity

Country welfare

$$w_i = \beta_i p_i (A_{-i} + \square_i e_i) - \tfrac{1}{2} c \alpha_i^2 e_i$$

Where: $\alpha = \frac{a}{e}$ denotes the fraction of emissions e that are abated, $B = \square p$ is total valuation of a unit of abatement a. and p is population

Threshold redefined

The threshold is now $\square$ and is *emission-weighted* (à la Kyoto): the treaty enters into force when ratifiers represent at least $k_{\min}$ of global emissions.

Nash & first-best (calibrated)

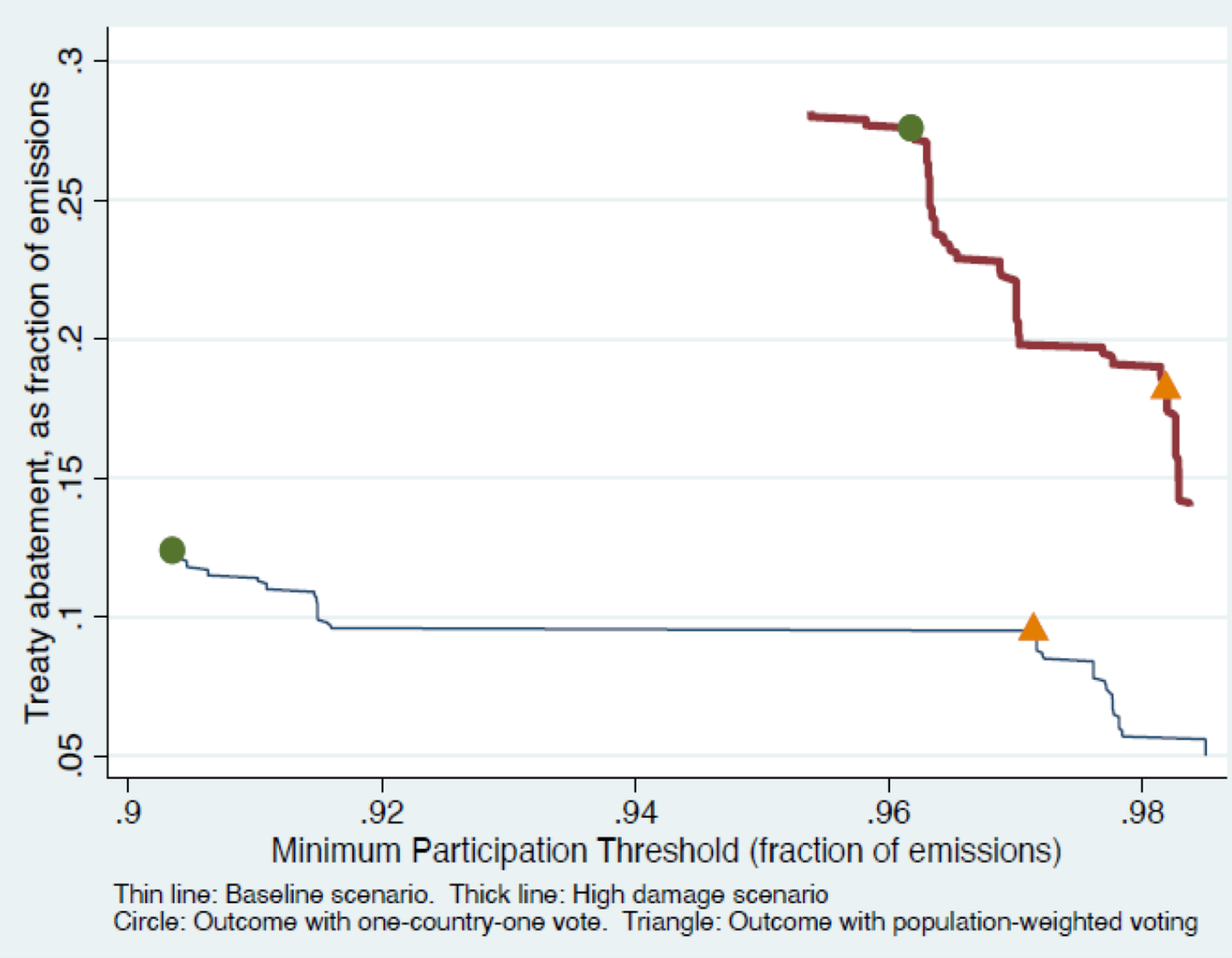
Baseline scenario

Nash **1.5%** of emissions · first-best **15.5%** — a ~10× gap.

High-damage scenario

Nash **3.2%** · first-best **31.1%** — same ratio, twice the level.

Threshold treaties cover > 90% of emissions



What the chart shows

Threshold treaties get close to first-best abatement.

Baseline. a^T up to 12.4% (first-best 15.5%).

High Damage. a^T up to 28.9% (first-best 31.1%).

Coverage. Threshold > 90% of global emissions in both scenarios.

China, USA and India are always members at the optimum.

Frontier shifts up & right under higher damages — more damage, more cooperation feasible.

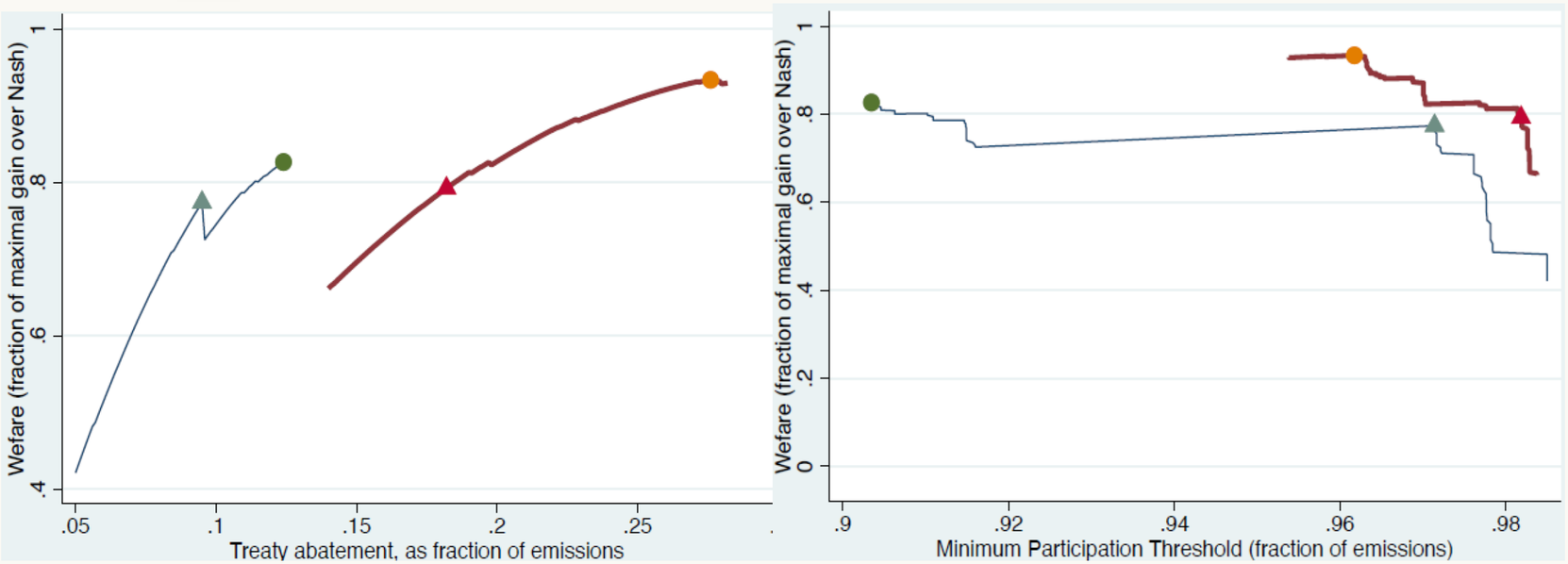
80–90% of first-best welfare gains

Baseline · UN rules

≈ 80% of the welfare gap between Nash and first-best, closed.

High damage · UN rules

> 90% closer still to the planner's ceiling when damages bite.



Thin line: Baseline scenario Thick line: High damage scenario

Circle: One-country one-vote Triangle: population weighted voting

Two further findings

i. Welfare dips when a major emitter exits

Baseline shows a visible dip when Russia drops out at $a^T \approx 9.5\%$.

ii. Population-weighted voting trades depth for breadth

"World Bank" rules give lower a^T but broader membership; gains 75–80%.

From theory to treaty design

Policy implications

i. **Thresholds are powerful institutional commitments**

Mitigate coordination failures that cripple public-good cooperation. Useful well beyond Kyoto and Paris.

ii. **Aim for the downward branch of the frontier**

Don't maximise breadth alone — combine ambitious targets with high but feasible thresholds.

iii. **The framework travels**

Biodiversity, pandemic preparedness, financial stability — anywhere a critical-mass clause matters.

Caveats & future research

Static model — no dynamics
Abates today's emissions; ignores changing technology, accumulated CO_2, and rising future damages.

No transfers, no enforcement
The threshold rule does heavy lifting alone. Side payments or sanctions could push further.

Country welfare ≠ national policy
Within-country distributional politics drive a wedge between model predictions and observed action.

No exit / renegotiation
Once-and-for-all ratification; real treaties allow withdrawal and updates.

Future: stochastic damages, learning, club benefits

Natural extensions to capture the full real-world complexity of treaty design.

Minimum-participation clauses are *institutional commitment devices*

> *A **concave ambition–participation frontier**, a **median-voter process that hits the constrained optimum**, and **80–90% of first-best welfare gains** under realistic calibration.*

Theory

A precise, non-monotonic tradeoff between depth and breadth.

Concavity guarantees an interior welfare-maximising treaty on the frontier.

Implementation

Sequential median voting matches the constrained optimum.

Democratic, decentralised governance need not sacrifice efficiency.

Quantitative

Threshold treaties cover > 90% of emissions; 80–90% of first-best gains.

Inclusive and ambitious cooperation is institutionally within reach.

Thank you · questions welcome