Validating CPAT: Baselines, Back-Casts and Model Comparisons

How far a finance ministry can trust the Climate Policy Assessment Tool

CPAT's published checks support it for ranges and for comparing policies, not for point forecasts. Its results turn on the price elasticities, and every check predates the 2026 bug fixes, so the checks should be rerun.

Stephen J. Stretton · writing in a personal capacity

Climate Fiscal Policy Brief No. 5.3 · October 2026

A carbon-pricing decision rests on a forecast, and CPAT's forecast has a wide error band

29%

G20 CO_2 cut in 2030 from a carbon tax rising to $75: CPAT's default result

7–39%

the same cut as the price elasticities move two standard deviations either way

231

bugs the 2026 rebuild found in the original model, 24 serious; every published check predates the fixes

If the response to the price is off by half, so are the revenue estimate and the gap to the pledge.

Source: CPAT methodology (2023), Parameter Sensitivity Analysis; CPAT-AI bug register, release 1.1.0 (2026).

Four tests answer four different questions, and passing one says little about the others

Base-year fit: does the model match the data in its starting year? Built in by calibration, so weak evidence.

Back-cast: started in 2000, does it reproduce what happened by 2017? The closest to an out-of-sample test.

Comparison: do other energy models, and studies of real carbon prices, give similar answers?

Replication: does independent code compute the same numbers? Verification, not validation, but a precondition.

Most comparisons are charts: the three with other energy models report no measure of fit.

Source: Working paper 5.3, Table 1; Sargent (2013); Wilson et al. (2021).

Calibration makes the base year right by construction and can carry a data break into every year

Emissions are scaled to the inventory. For Egypt, by 0.930: the 2021 inventory (211.05 Mt) over the model's 2022 estimate (226.97 Mt).

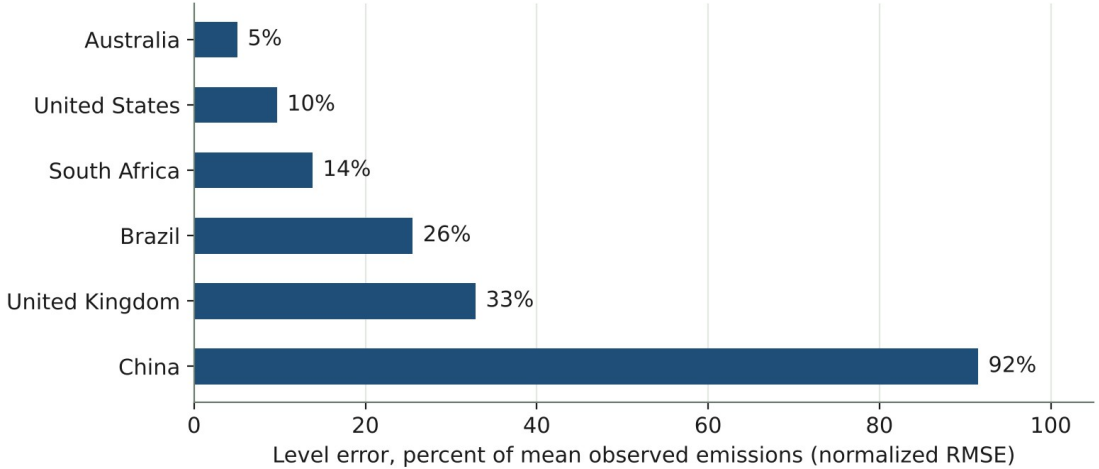
Then the inventory jumps 19.5% in 2022. The calibrated baseline runs 16%, 32% and 29% below the inventory in 2022, 2023 and 2024.

The CPAT-AI documentation says 16–29%: it misses 2023, where a separate fuel-use shock adds to the gap.

So compare the baseline with the inventory before using it; policy-against-baseline comparisons hide the gap.

Source: CPAT-AI model documentation, Emissions (2026); gaps recomputed in working paper 5.3, Section 4.

Run from 2000 with known world prices, CPAT's level error is 5% for Australia and 92% for China



Normalized RMSE of greenhouse gas emissions excluding land use, no-policy back-cast, 2000–2017.

Trend mostly right, level only where prices behave: China's projected domestic prices ran below observed ones.

A partial test: observed world prices were fed in, the hardest part to forecast.

Six large emitters, none a typical CPAT user and none low-income.

Source: CPAT methodology (2023), Hindcasting (CPAT 1.0pre_043); working paper 5.3, Table 2.

Against other energy models CPAT is of the same order; only power investment timing differs much

IEA STEPS (WEO 2021): world CO_2, primary energy and power mix of the same order; gaps on oil aggregation, Russia and nuclear.

Enerdata: the response of emissions to carbon prices of $0–120 has about the same slope.

World Bank power planning model (EPM), 34 countries: similar generation mix at about $40/t$CO_2$; EPM builds in lumps, CPAT spreads investment.

No measure of fit: "same order of magnitude" could mean 5% or 30%.

Source: CPAT methodology (2023), Model comparisons and Appendix H; working paper 5.3, Sections 4 and 6.

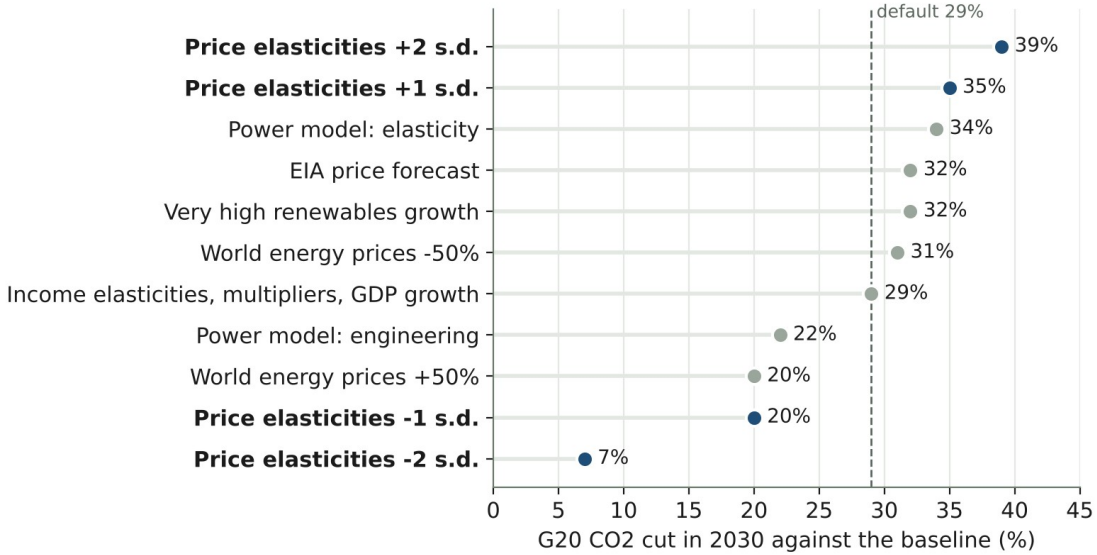
The evidence on real carbon prices is too wide to say CPAT over- or under-responds

Measure	Estimate	Source
CPAT, $10 price: average level gap, 2019–2023 / 2019–2035	−2.4% / −6.2%	CPAT methodology
CPAT, same, read as slower growth of emissions	0.7–0.8% a year	Author's calculation
Cross-country, 39 countries: per $10 of carbon price	0.5% a year (imprecise)	Rafaty et al. (2020)
Review of 37 studies: real schemes	0–2% a year	Green (2021)
Meta-analysis, 80 studies of 21 schemes	5–21% in total	Döbbeling-Hildebrandt et al. (2024)

CPAT is of the same order as the most comparable estimate. Robust finding: power responds most and transport least, in CPAT and in the evidence.

Source: Working paper 5.3, Section 6.2 and Table 3. Conversion assumes the level gap builds linearly.

The price elasticities move the G20 cut from 29% to anywhere between 7% and 39%



Power model next: 22% (engineering) to 34% (elasticity).

Income elasticities, multipliers, GDP growth: almost no effect on the cut.

Gaps in the evidence: household electricity (−0.1 estimated vs −0.4 in CPAT) and services; gasoline (−0.61) sits inside the literature's range.

Source: CPAT methodology (2023), Parameter Sensitivity Analysis and Panel estimation; Brons et al. (2008); working paper 5.3, Table 4.

Independent code reproduces CPAT exactly, and the rebuild found the bugs the checks were run on

Legacy blocks reproduced in Python: road model to 3.8×10^{-14}, power prices to 5.6×10^{-17}, power model to 10^{-13}.

Coded CPAT = Excel-AI: 7,458 output codes per scenario within 10^{-9}; one country in under a second.

Against legacy, with bugs fixed: all 15 CO_2 rows within 5% in 2022; median CO_2 gap 12% by 2030.

Six data tables are proprietary: outsiders can rerun the method, not every country.

Source: CPAT-AI model and Coded CPAT documentation, release 1.1.0 (2026); working paper 5.3, Table 5.

Read CPAT as a range, compare policies with it, and make validation a routine test

How to read a CPAT number

A 29% cut means about 20 to 35%: elasticities and power model.

Comparisons between policies beat levels; baseline errors cancel.

Check the elasticities where a result hinges on household electricity, services or road fuels.

Check the baseline against the national inventory.

What to do

Ministries: ask for ranges and a measure of fit with every result.

Country teams: state the power model, elasticities and data vintage.

CPAT maintainers: rerun back-casts, comparisons and sensitivity on the corrected model, every country, every release.

Researchers: test CPAT against real carbon taxes, per dollar.