

THE EU CARBON PRICE TWO YEARS AHEAD

A TESTED DISTRIBUTION BUILT FROM THE LAW, THE MARKET AND POWER-SECTOR
DISPATCH — THE FORWARD-PROJECTION FRAMEWORK PROMISED IN WORKING PAPER 02

Peter Keutgens, FIA

Independent Research Note · Working Paper 05 / 2026 · Carbon Risk & Capital Allocation Series

September 2026 · Data version 2026-09-21b (content id d5dd0c75d073) · package svar_generator 0.4.1

Companion to Working Paper 01 / 2026 (carbon as a capital risk factor) and Working Paper 02 / 2026 (the componentised cap). Comments and corrections welcome; the revision record and the claim register are available on request.

EXECUTIVE SUMMARY

A finance director who carries a carbon cost — an allowance surrender, a certificate liability at the border, a covenant on a carbon-exposed loan, a scenario line in an ORSA — has one problem in four forms: what the EU allowance price will be over the horizon on which that provision, covenant or scenario is set, with what uncertainty, and what it does to the emissions the cost is charged on. This paper's horizon is twenty-four months, the reach over which the market's own history can be tested; beyond it the paper says what it cannot say. Working Paper 01 argued that under a quantity-constrained system the price is a capital risk factor whose distribution, not forecast, is what a decision needs; Working Paper 02 reconstructed the supply side from the legal record and promised the forward-projection framework that would use it. This is that paper. It builds the distribution of the allowance price and of power-sector emissions two years ahead, tests it at fifty-six past month-ends against what then happened, and runs every morning on the platform that holds the cap. Model results are generated from versioned data with a version stamp, while external numerical and legal inputs are separately sourced; both are read against their source by a generated register.

Six findings.

1. Supply is written in law, and it does not forecast the price. Nothing the law and the registry know at a given month — the published surplus, the reserve's coming intake, the auction calendar, recent auction cover — improves a forecast of the price one or two years out, when the model is re-run at 52 past month-ends on the information available then and scored against a random walk. None of the four signals met the rule set before the test. That is what one would expect of a market in which about a billion tonnes are banked — what is known is already in the price — though the test establishes the absence of skill, not the reason. The law's contribution to a near-term distribution is the quantity ledger, the rules that act in the tails, and the surprises when the law changes.

2. The market is rarely surprised by carbon policy, and less often than it used to be. A register of 202 regulatory events since 2005, each tied to a stored source document, identifies policy shocks clearly in 2008–2014, when the reform was contested, and only weakly from 2015, when the process settled into

one designed to anticipate rather than announce: the events still come and the price still moves on the day, but the moves are half the size and explain little of a month's change. A surprise that lifts the price by one per cent leaves about three-quarters of it in place three years later; whether the effect is permanent the evidence cannot say.

3. The carbon price is associated with power-sector emissions through dispatch, and the association has a threshold. On 1,519 trading days, each euro by which the allowance sits above the coal-to-gas switching price is associated with about 0.15 percentage points of coal-plus-gas generation moving from coal to gas — roughly 0.11 per cent of those emissions per euro — and below the threshold, with no measurable change. Read through that estimated association, a ten-per-cent price surprise corresponds to a 0.3–0.5 per cent fall in power emissions in a year like the last one and to no estimated fall in a year like 2022. The estimated emissions consequence of a carbon price is state-dependent, and the state is set by the fuel markets.

4. Over the tested sample and horizons, the model did not forecast the monthly allowance price more accurately than a random walk — and the honest model is built to say so. A monthly system estimated on 2019–2026 loses to a driftless random walk by a factor of four at one year, because it learned the 2019–2022 quadrupling as its normal; shrunk toward a random walk, in differences, it recovers near-parity and no more, at every lead and in every year of the sample. The model's job was never to guess the direction of the price. Its value is in what the random-walk benchmark — which has a width of its own, and over-covered in the same way — does not give: tails shaped by the policy jumps and fuel shocks actually observed, the switching response of finding 3 and the emissions line of finding 6.

5. The distribution is deliberately wide, and the record so far says wider than needed. Since 2023 the price has stayed within a range while the fan carries 2022 in its width, so the central 80 per cent band contained the realised price almost every time from six months out — as the random walk's did — and was about three times as wide as the errors it had to cover. Narrowing the fan to four calm years was tried in a limited form and refused in full: it would break at the next 2022-sized move. The fan is kept as a conservative risk envelope, by decision, and the production record will show whether the calm holds.

6. The emissions line beat the random walk while emissions were falling, and matched it since. Pooled, the emissions fan is 31 and 49 per cent closer to the outcome than the random walk at one and two years. Cut by year, the skill is concentrated in forecasts made during the 2022–23 decline; on the 2024–25 dates it matches the random walk and under-covers. Whether the skill is the equation's or the episode's, the forward record will settle.

Against the NGFS paths, the distance is coverage first and price second. The NGFS carbon price is an economy-wide shadow price; the allowance reaches 38 per cent of EU emissions and, net of free allocation, 22 per cent. Weighted by the tonnes that pay, the EU's effective carbon price today is €12.5 per tonne in 2010 money against a Net Zero 2050 path at €26, and the allowance would have to be €169 today and €286 by mid-2028 for the two to coincide. Extending pricing to road transport and buildings — the second trading system, scheduled to become fully operational in 2028 — if priced like the existing system, would close most of that distance in one step — an illustration with strong conditions attached, one model's EU-15 path read against EU27 shares, and the clearest statement of where the gap comes from.

For the finance director this gives a carbon-cost line with a stated interval and a reported historical coverage, an emissions line for power-sector exposure and a price distribution to apply elsewhere,

and a scorecard that gives, horizon by horizon, the probability that the EU's coverage-adjusted carbon price reaches a scenario's carbon-price path — not the probability of the climate outcome — with its conditions attached; §9 reads it for the investor, the lender, the insurer and the importer. The long horizon is not modelled here: §10 sets out how the structural layer will join this one, and the long-term scenarios remain the NGFS paths as given.

1. WHAT THIS PAPER DELIVERS, AND WHERE IT SITS IN THE SERIES

Working Paper 01 (February 2026) set out a framework in which the carbon cost of a long-lived asset is a distribution across thousands of policy and market paths — a structural, long-horizon and necessarily assumption-driven model. Working Paper 02 (June–July 2026) reconstructed the supply side year by year from Commission decisions and registry data, and promised the forward-projection framework that would use it. This is that paper, with one deliberate change of horizon: rather than project the structural model forward on assumptions, it measures what can be measured — how the market reacts to policy, how the price reaches emissions, how price and fuels move month to month — and states what that implies for the next twenty-four months, with the whole construction tested against the realised past. Beyond two years the market's own history says little, and the long-horizon structural layer waits for the paper that follows (§10).

Two commitments run through everything below. First, **every layer was tested against a rule recorded before the run, on an as-of backtest against a stated benchmark**; failures are reported as results, and a layer that failed is retained only where the risk-management judgement for keeping it is made explicit (§7.1 is the case). "As-of" means the model is re-estimated at each past month-end using only what was known then, run forward, and scored on what happened. Second, **the work runs on a platform, not in a spreadsheet**: the cap components, the event register, the daily prices, the dispatch data, the forward curves and the NGFS paths live in one database with their provenance; the generator is a versioned package the platform runs every morning, storing each fan with the data version, code version and parameter set that produced it. The backtest in §7 is therefore the beginning of a record, not the end of one.

Scope. The EUA price fan applies across the sectors the EU ETS covers — power, industry, aviation and, from 2024, maritime — because the allowance is one fungible instrument. The emissions-response model in this paper is confined to coal- and gas-fired electricity generation. Aviation is not yet included: the platform's flight feed had too short a history when this model was specified to estimate and backtest an aviation response, and aviation will be added later as a separate allowance-demand module. No aviation emissions response is inferred in this paper, and no industrial one either: readers with those exposures apply the price distribution to their own emissions estimates.

Section 2 gives the design in a page. Sections 3–5 present the three measured pieces — the legal supply, the policy surprise, the switching hinge. Section 6 assembles the generator and Section 7 tests it. Section 8 reads the result against the NGFS scenarios. Section 9 is for the reader with a balance sheet. Section 10 states limitations and next steps. A closing note says what further material is available.

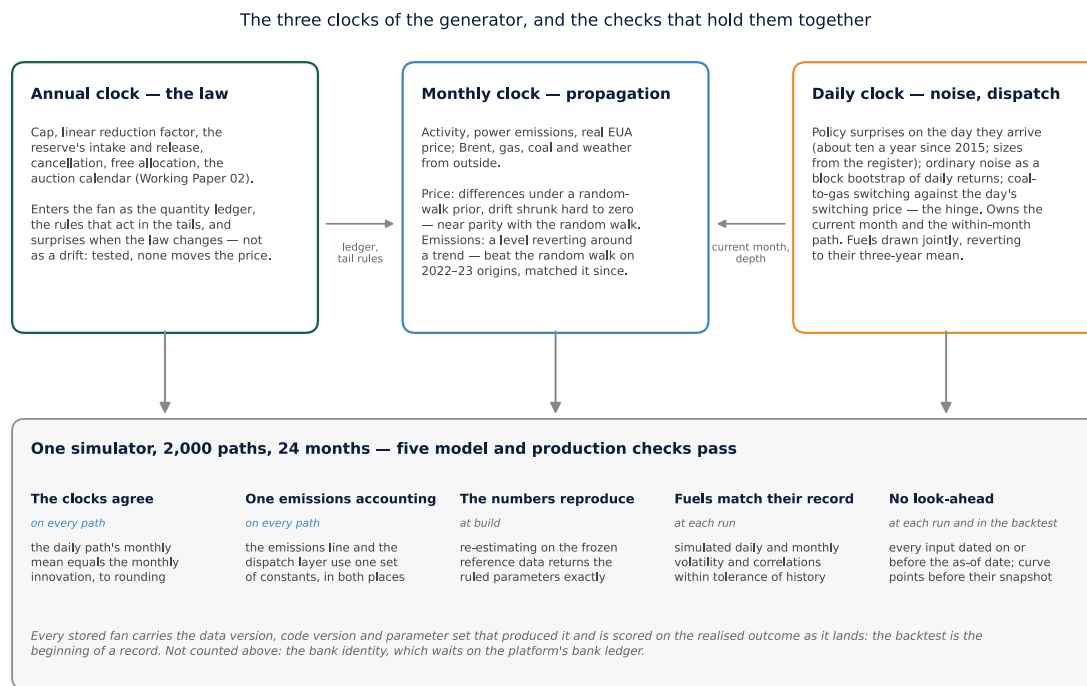
2. THE DESIGN: THREE CLOCKS, ONE STATE, AND A DISCIPLINE

The EU ETS operates on three clocks, and the model keeps them separate.

The annual clock is the law. The cap for each year, its linear reduction, the Market Stability Reserve's intake and release, the cancellation rule, free allocation and the auction calendar are set by legal act and known in advance, with a small number of decision points each year (the Commission's publication of the surplus, due by 1 June and issued in late May; the reserve's intake window turning over on 1 September; the surrender deadline of 30 September — Figure 3). Working Paper 02 is the reconstruction of this clock. Nothing on it moves faster than a year, and — the finding of §3 — nothing on it that is known at a given date carries information the price has not already absorbed.

The monthly clock is propagation. Activity, power-sector emissions and the allowance price, with the three fuels (Brent, gas, coal) and weather from outside, form a small system whose month-to-month co-movements can be estimated. This is the clock at which one asks how a fuel shock or an activity shock works through to price and emissions over a year or two.

The daily clock is noise and dispatch. Policy surprises arrive on specific days and are measured on those days (§4). System dispatch shifts generation between coal- and gas-fired units each day on that day's prices (§5). Both are invisible at monthly frequency: a monthly average smears a one-day jump across two months, and a monthly model of dispatch cannot see a threshold crossed and re-crossed within the month.



Rolling two-year trading-horizon model — not a strategic projection; the reader's exposure is usually longer than the fan's reach.

spec §2.0 • package 0.4.1 • gate G4c 2026-09-25/26 • render_wp05.py E01

Figure 1 — The design in one picture: the annual, monthly and daily clocks, what each supplies to the simulator, and the five model and production checks that pass. Two of them hold on every simulated path — the clock bridge and the single emissions accounting; the other three are build-, run- or backtest-level checks (reproduction of the ruled parameters, the fuels' match to their record, no look-ahead). A sixth, the bank identity, waits on the platform's bank ledger and is not counted.

The generator (§6) composes the three. Monthly innovations from the estimated system carry the price and the fuels from month to month; inside each month the daily layer supplies the policy jumps, the ordinary daily noise and the dispatch response, and owns the current month, where it knows the price at the last close rather than the month's average; the annual layer supplies the quantity ledger against which emissions are read and the rules that act in the tails. Two consistency checks hold on every simulated path — the daily path's monthly mean equals the monthly innovation, and the emissions identity holds with one set of dispatch constants in both places — and three more at build, run or backtest level: the ruled parameters reproduce, the fuel moments match the record they were cut from, and no input is dated after the run's as-of date (Figure 1).

3. SUPPLY IS WRITTEN IN LAW — AND IT DOES NOT FORECAST THE PRICE

Working Paper 02 established what the cap actually was, year by year: not a smooth decline but a componentised object with nine material mechanisms most descriptions omit. Figure 2 is that reconstruction in one picture, and it is the starting point for everything in this paper.

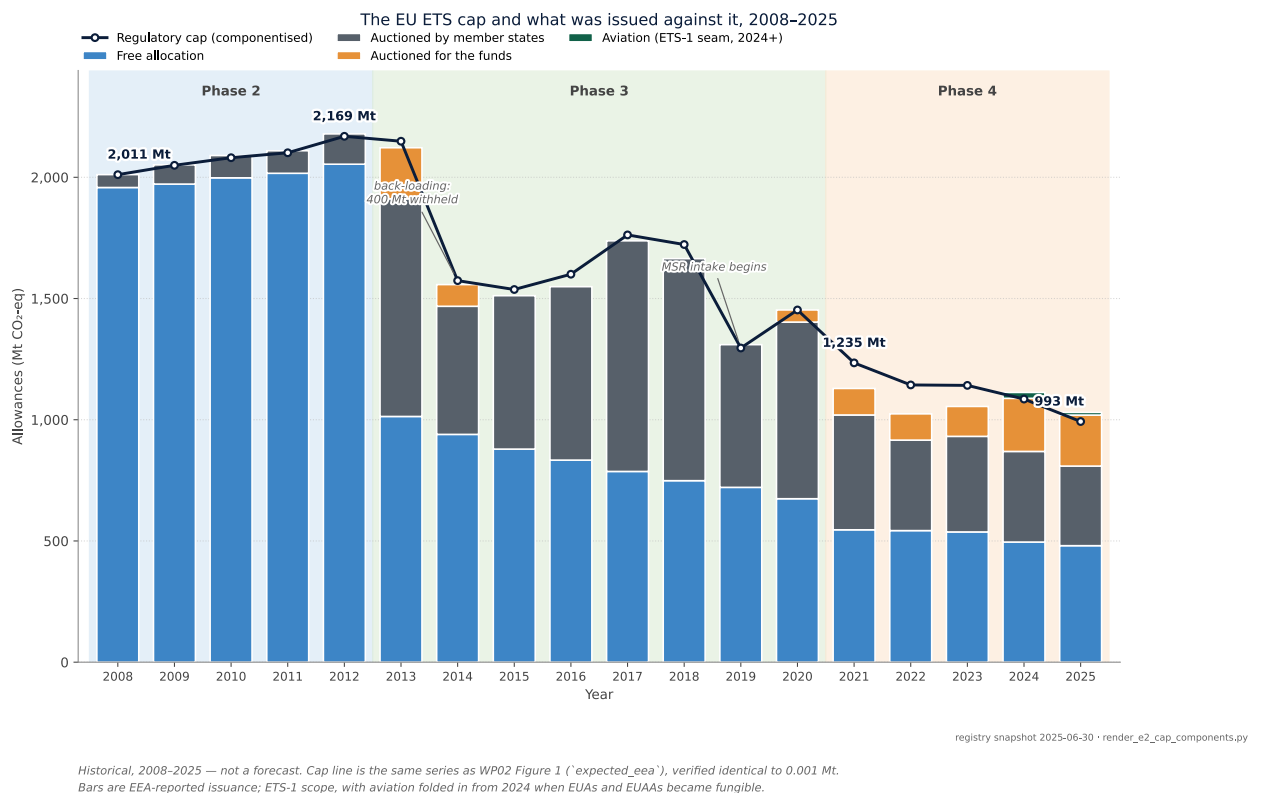


Figure 2 — The cap (line) is the componentised regulatory cap of Working Paper 02, verified identical to that paper's series; the bars are what the registry reports as issued each year — free allocation, member-state auctions, auctions for the funds, and from 2024 the aviation quantities folded into the single pool. The two interventions that moved supply off its statutory path, back-loading in 2014 and the reserve's intake from 2019, are named where they act. Platform figure; registry snapshot 30 June 2025.

Two things in Figure 2 matter here. The first is that the legal supply path is knowable in advance to a degree without parallel in other commodity markets. The cap for 2027 is a number in a Commission decision; the reserve's intake for the twelve months from September is a formula applied to a surplus the Commission publishes each spring; the auction calendar is fixed for the year. Figure 3 shows what that looks like inside a single year — the weekly auction supply, and the three dates the law fixes on it.

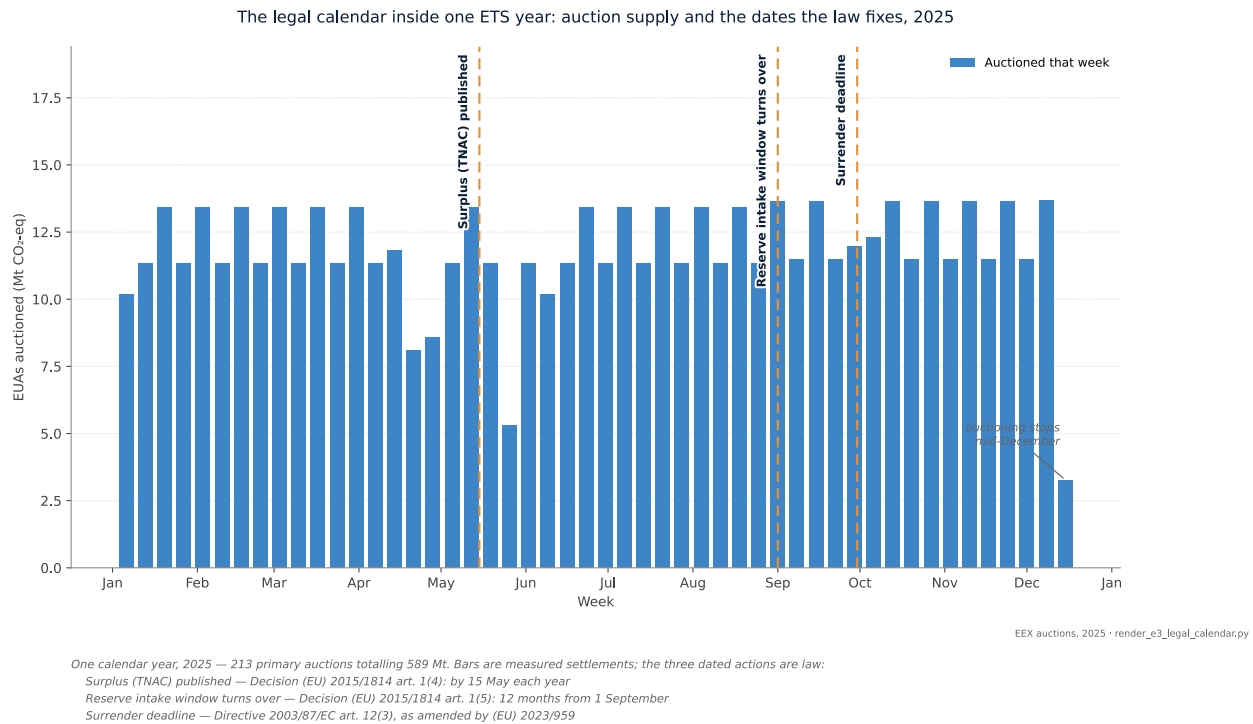


Figure 3 — One calendar year, 2025: 213 primary auctions totalling 589 Mt, week by week, with the three dated legal actions — the surplus (TNAC) publication in May, the reserve's intake window turning over on 1 September, the surrender deadline at the end of September — and the mid-December pause. These are the things a forecaster knows in advance; §3 tests whether knowing them helps. Platform figure; EEX auction data.

The second thing, established in this paper, is that **none of the four tested policy signals added out-of-sample predictive skill** — the reading being that the market has already priced what the law knows. Four signals, each known at the origin month or readable from Figure 3's calendar — the published surplus and its year-on-year change; the reserve's coming intake as a share of the cap; the auction volume for the coming months against its trailing year; and the auction cover ratio — were added one at a time as drift terms to the monthly price equation and tested as-of across 52 month-ends since January 2022, under a rule recorded first: a signal stays if it lowers the twelve- and twenty-four-month forecast error against the random walk by more than five per cent without costing coverage. None does. The surplus signal is the most interesting failure, because it has the sign economics expects — a shrinking bank goes with a rising price — but the relation is weak by every measure: a t-statistic of 1.2, a coefficient of 0.001 once shrunk toward the no-information prior, and a change in the forecast-error ratio of less than 0.01. The largest change any signal produced was 0.04, below the rule's 0.05.

This is not a failure of the annual layer; it is what a market for a bankable asset with a bank of around a billion tonnes should do. If a rule known in May will tighten supply from September, an allowance bought in May is worth what it will be worth in September less the cost of carry, and the price moves in May or before. The annual layer therefore enters the generator in three other ways: as the quantity ledger the emissions line and the scorecard are read against; as the rules that bend the tails — the reserve's thresholds and the Article 29a release of 75 million allowances when the six-month average price exceeds 2.4 times the two-year reference, which act at price levels the 80 per cent band does not reach; and, decisively, through *changes* in the law, which are surprises, and the subject of the next section.

One book-keeping item belongs on the record, in proportion. Figure 2 shows that the cap reconstruction reconciles to the registry year by year, to within the interventions it names. A second, narrower identity — this year's published surplus less last year's should equal the allowances issued less the emissions surrendered — cannot yet be closed from the cap components alone: a cap-based proxy leaves residuals of between about 40 and 210 million tonnes a year over 2019–2025, in both directions, which measure what the proxy omits (free allocation as issued rather than as entitled; the aviation quantities, which enter the surplus calculation only from 2024 and count issuance and emissions only from 1 January 2024, per Working Paper 02 §6.6; the unallocated and reserve quantities; cancellations), not allowances that are missing. The platform is building the ledger that closes it. Nothing in this paper depends on it: the model takes the published surplus as data, and its bank check is reported as not run rather than as passed until the ledger lands.

4. WHAT THE MARKET HEARS: THE POLICY-SURPRISE SERIES

4.1 The construction

The method is borrowed from monetary economics, where a central bank announces a decision at a scheduled moment and the change in an asset price in a narrow window around it measures the part that was not expected. Känzig (2023, *American Economic Review*) carried this design into the EU carbon market with a chronology of regulatory events and the allowance price's reaction on each day.

Our register was built without knowledge of his chronology — 202 eligible events on 194 trading days from 2005 to 2026, each tied to a source document in the platform's store, dated by the day the information reached the market, with a once-per-day rule where two events coincide — and the surprise is measured as the allowance's log return on the event day against its December-benchmark contract, normalised by the trailing volatility, a normalisation chosen without reference to his. Two construction lessons cost several weeks and are worth recording. Phase-1 allowances were not bankable and traded down to one cent in 2007, so a percentage return on a contract at €0.10 is tick noise, not information: nine early events were quarantined for this reason, and Figure 4 draws them hollow rather than dropping them, so the exclusion is visible. And two overlapping price sources for 2021–2025 double-counted every trading day in a first build; every daily price now passes through one deduplication rule.

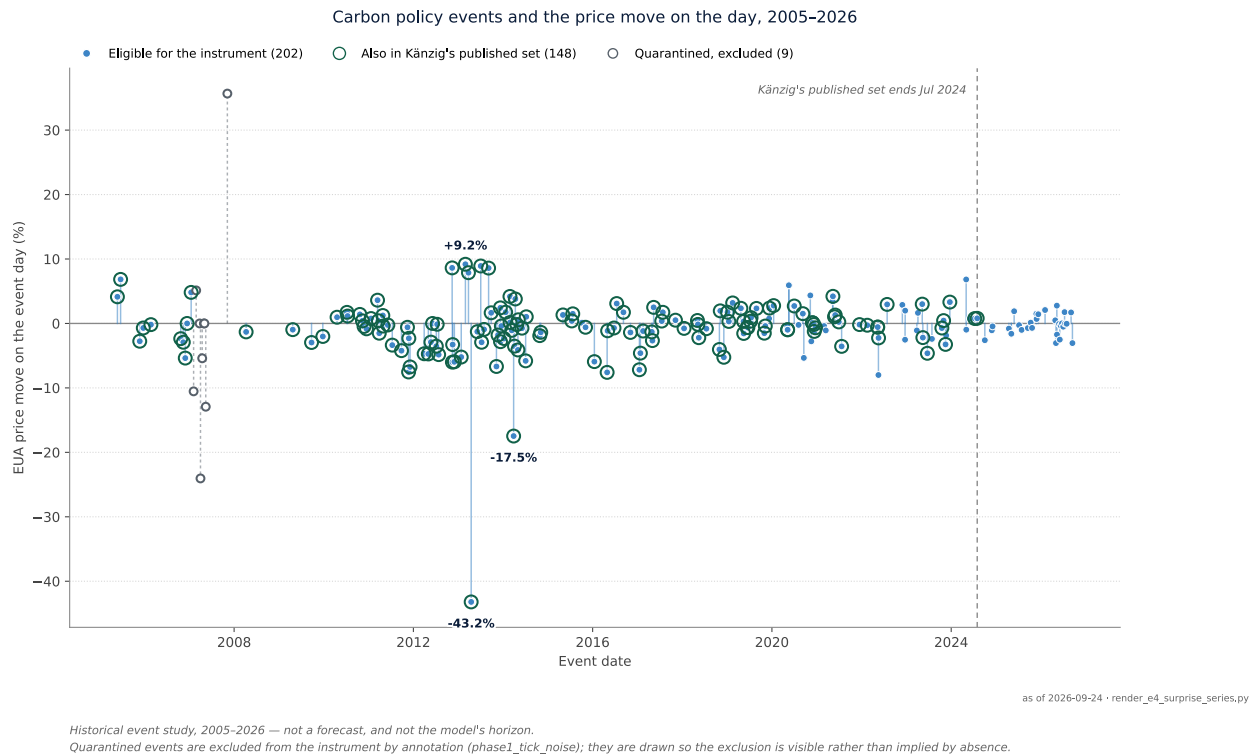


Figure 4 — Every event in the register and the allowance's move on the day: 202 eligible events (filled), the 148 that Känzig's published set also contains (ringed), and the nine phase-1 events quarantined as tick noise (hollow, drawn so the exclusion is visible). The largest moves are the contested years — the April 2013 rejection of back-loading at -43 per cent — and the series thins visibly after 2015. Platform figure, as of 24 September 2026.

The register was then checked against the benchmark, not confirmed by it. On the 114 event days the two chronologies share, our normalised surprise correlates with his published series at 0.81. On the years his paper does not cover, the events we found independently through the Commission's archive recover 28 of the 45 days in the later extension of his series, at a correlation of 0.85; seventeen of his post-2019 events we did not find, a recall gap on the platform's queue. The overlap statistics live in the platform's tables, not in this paper's result files, and are listed as such in the claim register.

4.2 Where the surprises are — and the reading

The series is meant to be used as an instrument: a source of variation in the price that is intended to come from policy alone. An instrument has to satisfy two things. That it is exogenous — that the surprise moves the price for policy reasons and not because something else moved both — is an assumption maintained on the construction of the events and on the evidence around them; it cannot be tested by a statistic, and §4.1 and the register are the case for it, not a proof of it. That it is relevant — that it accounts for enough of the monthly price change to be usable — is tested by its first stage, and here the result is sharp (Figure 5).

Measured on end-of-month prices in a monthly system from 2008, the first-stage F-statistic is 43 on the full sample, 25 on 2008–2014, and 1.9 on 2015–2026. The identifying power is concentrated in the years the reform was contested: phase 2, when the market was learning whether the EU would act on the surplus at all, and the back-loading fight of 2013–2014, when the process broke visibly (a plenary rejecting what the committee had agreed in April 2013; a reserve proposal markedly more aggressive than trailed in 2014). It thins from 2015, the year the reserve was adopted, and stays thin through the phase-4 revision and Fit for 55 — the largest tightenings in the system's history. Känzig's own series, extended to 2024 by its authors, shows the same pattern on the same first stage.

Where the policy surprises identify the price — and why the price must be measured at month-end

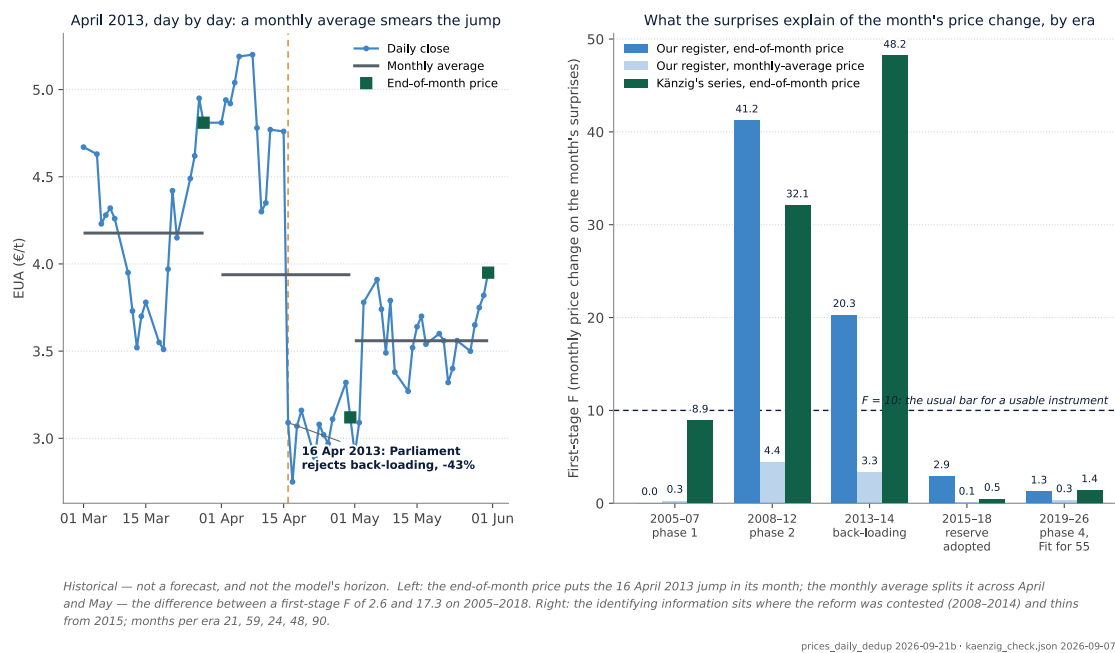


Figure 5 — Left: April 2013 day by day. The end-of-month price puts the 16 April jump in its month; the monthly average splits it between April and May. Right: the first-stage F-statistic by era — how much of a month's price change the month's surprises account for — for our register on end-of-month prices, our register on monthly averages, and Känzig's series. The usual bar for a usable instrument is 10. Months per era: 21, 59, 24, 48, 90.

The first thing to say is that the instrument is weak at monthly frequency after 2015 — event-day moves continue, at half their earlier size (\$4.3), but they no longer account for much of the monthly price change; the interpretation — that a process which trails a proposal for months, negotiates Council and Parliament positions in public, and publishes months after the plenary is one built to iron out disagreement before any moment arrives — is a reading, offered as such. A market watching that process should be surprised rarely, and when it is, the surprise should come from the process failing to work as designed, or from a crisis that forces policy outside its cycle. The 2021–22 energy crisis produced exactly one such episode in the register, the REPowerEU sale of allowances from the reserve, proposed nine months after the gas price spike. That single episode is the only observation we have of how the EU's legal construct responds to a serious energy shock, and it is why the generator treats the policy reaction to such shocks as an open question rather than the surprise distribution as a given.

A methodological point that cost a week and changes the result more than six-fold: the price must be measured at the end of the month, not as the month's average. A monthly average smears a single-day jump across two months' changes; the end-of-month price attributes it to its month. On 2005–2018 with the same events, the first stage moves from 2.6 to 17.3 on that choice alone (Figure 5, left).

4.3 How long the reaction lasts, and what it does

Two questions follow. Does the market's first reaction hold? And what does a policy surprise do to the price over years?

On the first (Figure 6): for each event day, the price change from that day to each of the next twenty-four trading days is regressed on the event-day move itself, across all events. The result is a share — how much of the event-day move is still in the price h days later — which is one on the day, stays at one if the market keeps the move, and falls toward zero if it reverses; ordinary non-event days put

through the same regression revert to 0.85 within a month, so persistence is measured against that placebo. In the contested era two-thirds of the event-day move is gone within a month, and part of that is the policy itself reversing — the April 2013 rejection was overturned in July. Since 2015 the events still come and the price still moves on the day — the average move is 1.9 per cent against 3.7 before — and those smaller moves are kept: the share sits at one after a month, on a band wide enough to say little more. What weakens after 2015 is not the market's reaction on the day but the size of the surprises, and with it the instrument's power at monthly frequency.

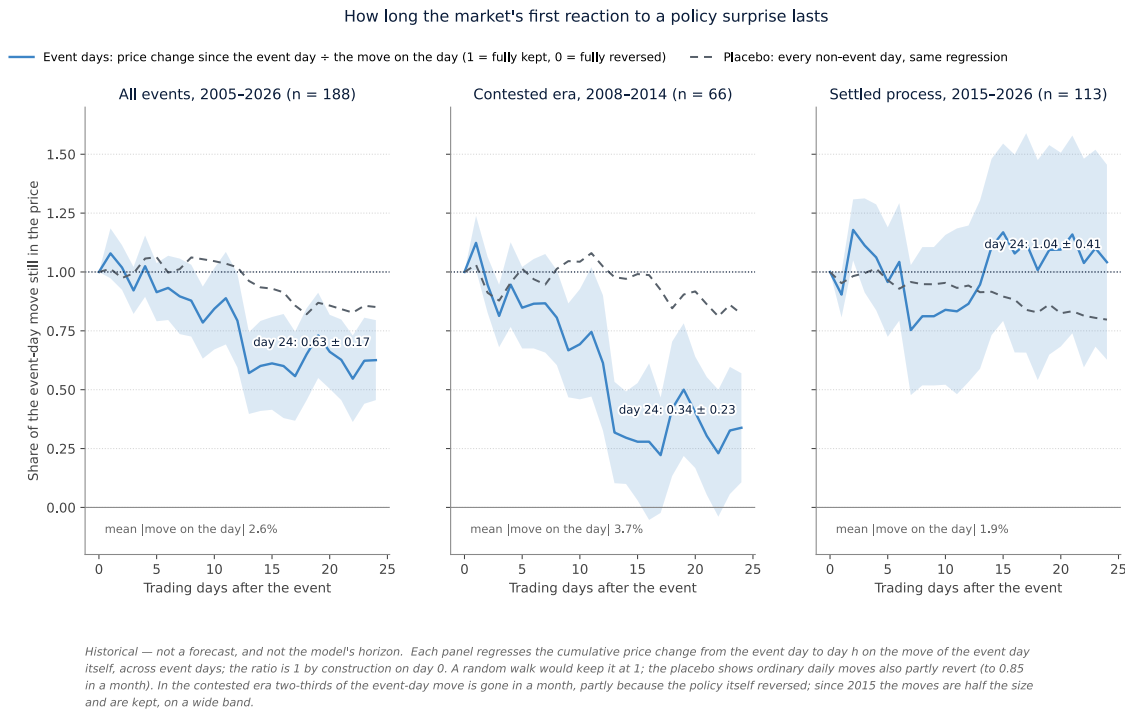


Figure 6 — The share of the event-day move still in the price, by trading day after the event, with a one-standard-error band; the dashed line is the same regression on every non-event day. All events; the contested era; the settled process since 2015. Mean absolute move on the day: 2.6, 3.7 and 1.9 per cent.

On the second (Figure 7): in a monthly system of energy prices, ETS-weighted industrial activity and the real end-of-month allowance price from February 2008 — Känzig's design, our instrument — the question is what a policy surprise does to each of the three over the following three years. Two estimators answer it. A vector autoregression fits the month-to-month dynamics once and iterates them forward, giving a smooth path whose bands inherit whatever the model gets wrong about persistence (hence the bias correction); a local projection regresses the change at each horizon directly on the surprise, so nothing compounds and the uncertainty is estimated where it arises. When the two agree, as they do here once later surprises in the window are held fixed, that is worth more than either alone. A surprise that lifts the price by one per cent leaves 0.79 per cent of it in place after three years on the vector-autoregression estimate and 0.76 on the local projection; the latter's standard error at three years is 0.44 and its weak-instrument-robust 95 per cent set runs from -0.01 to $+1.75$. **The estimated long-run response is large and imprecise: the data do not reject full persistence, and they do not establish it.** Energy prices rise slowly, by about 0.1 per cent at three years; the activity response is small and not robust across estimators. On the day of the shock there is no statistically detectable contemporaneous response in activity or energy prices. That is consistent with the exclusion restriction, but it does not establish it: the restriction remains a maintained assumption, and the same-day evidence says only that the data do not contradict it.

A carbon-policy surprise in the long-sample system, February 2008 to June 2026: the price move persists

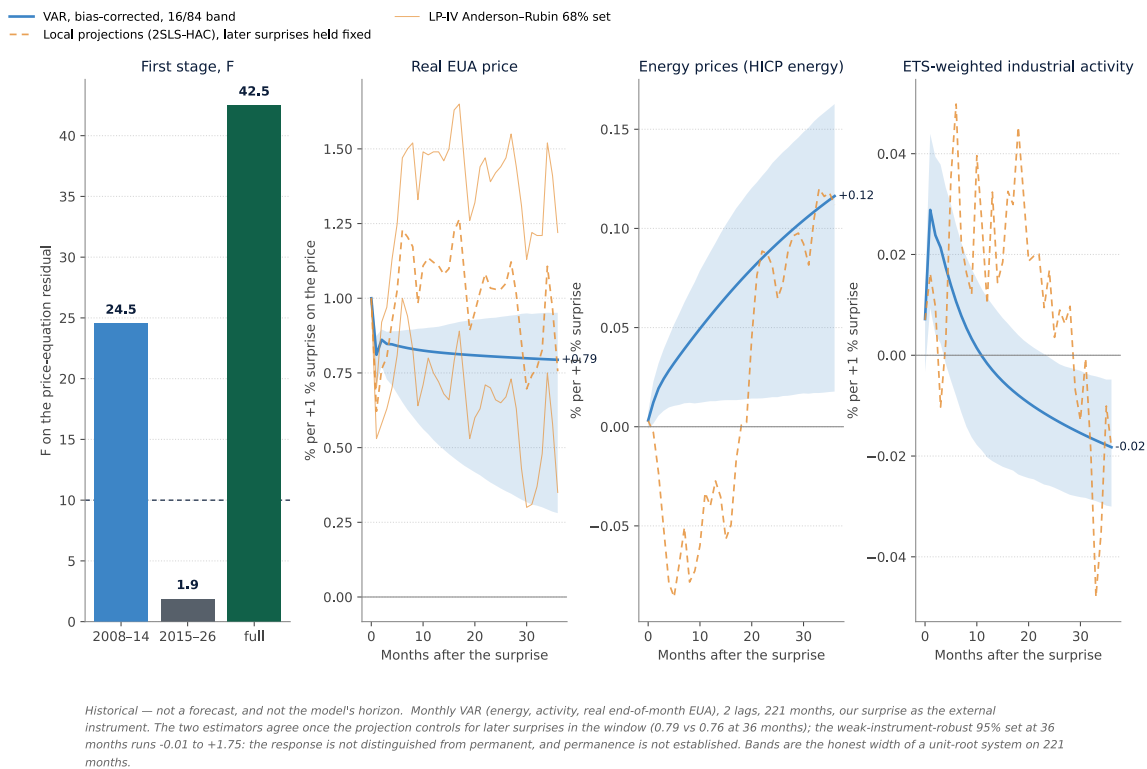


Figure 7 — Left: the instrument's first stage by era. Right three panels: the response of the real allowance price, energy prices and ETS-weighted industrial activity to a policy surprise that lifts the allowance by one per cent, month by month for three years — the vector-autoregression path with its 16/84 band, the local-projection path (two-stage least squares, later surprises held fixed) with its Anderson–Rubin 68 per cent set. The price response is large and imprecise; energy prices rise slowly; activity barely moves.

4.4 What monthly data cannot do

A monthly system of activity, emissions and price on 2019–2026 — the years the platform's fuel and dispatch data cover — cannot be identified from its own co-movements. Such a system has two parts: the dynamics — how each variable's past months predict this month — which can be estimated on 2019–2026 and are what the monthly engine of §6 uses to propagate a shock; and the within-month surprises, from which "the policy shock" would have to be identified. On 2019–2026 the price and emissions surprises are nearly uncorrelated, so the system carries almost no contemporaneous information to identify with; a sign-restricted identification returns essentially its own prior, and the daily instrument's first stage on these years alone is 1.3. That determines the design: what the model knows about causation comes from the long sample and the daily data, not from the monthly system, whose job is propagation.

5. FROM PRICE TO EMISSIONS: THE SWITCHING HINGE

Scope of this section. The price fan of §6 applies to every ETS-covered tonne; the emissions response estimated here covers coal- and gas-fired electricity generation only. Industrial process emissions are outside it. Aviation is outside it because the platform's OpenSky flight feed, live since June 2026, had 84 days of history when the model was specified — too short to estimate and backtest a response; aviation will be added as a separate allowance-demand module once the history allows, linking flight

activity and fuel costs to allowance demand. Nothing in this section should be read as an estimate of a steelmaker's or an airline's abatement.

The allowance price reaches emissions in the near term through one mechanism above all others: the daily choice between coal and gas in power generation. A gas plant emits roughly half the CO₂ of a coal plant per unit of electricity, so when the allowance price is high enough to make gas the cheaper fuel including carbon, dispatch switches and emissions fall. The breakeven — the switching price — depends on the gas price, the coal price and the two plants' efficiencies, and moves every day.

Figure 8 shows the two prices day by day since 2019 and the coal share that answers them. Where the allowance sits above the switching price — the shaded stretches of 2019–2020, 2023 and early 2024, the spring of 2026 — the coal share falls; through 2022, when gas took the switching price above €1,000, the allowance was nowhere near it and the coal share did what the winter told it to.

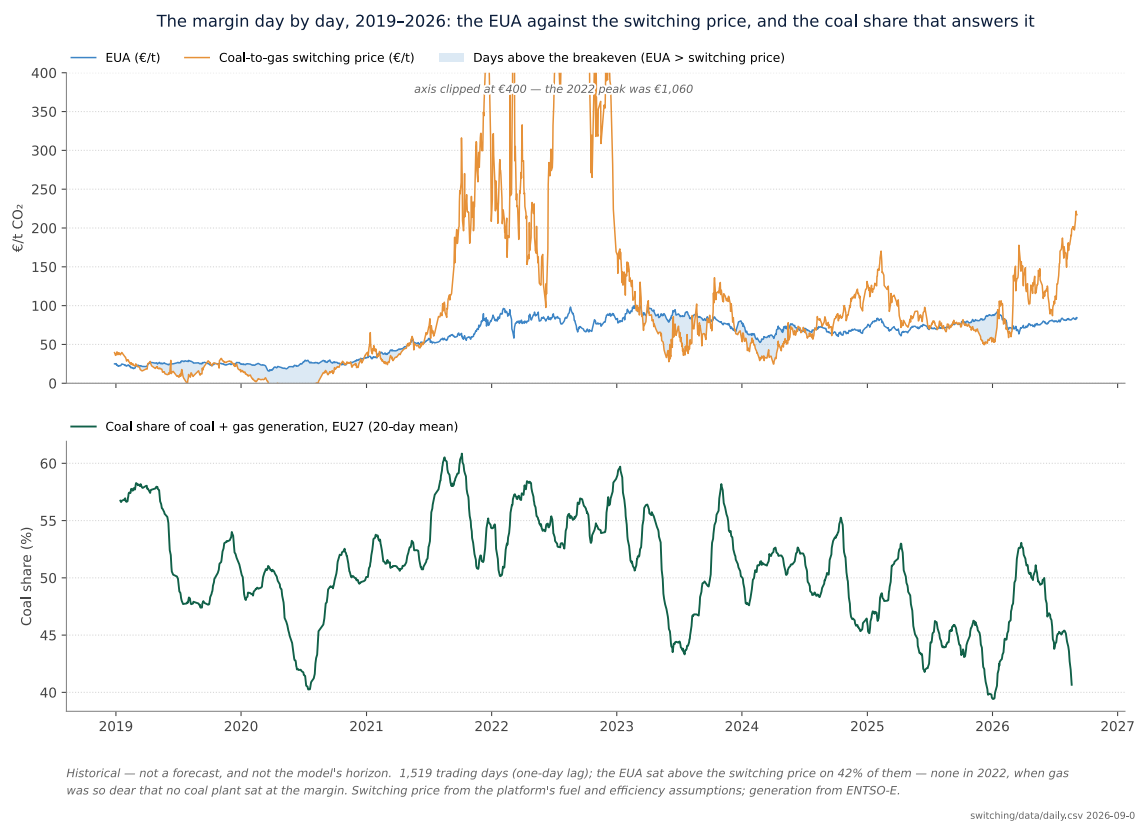
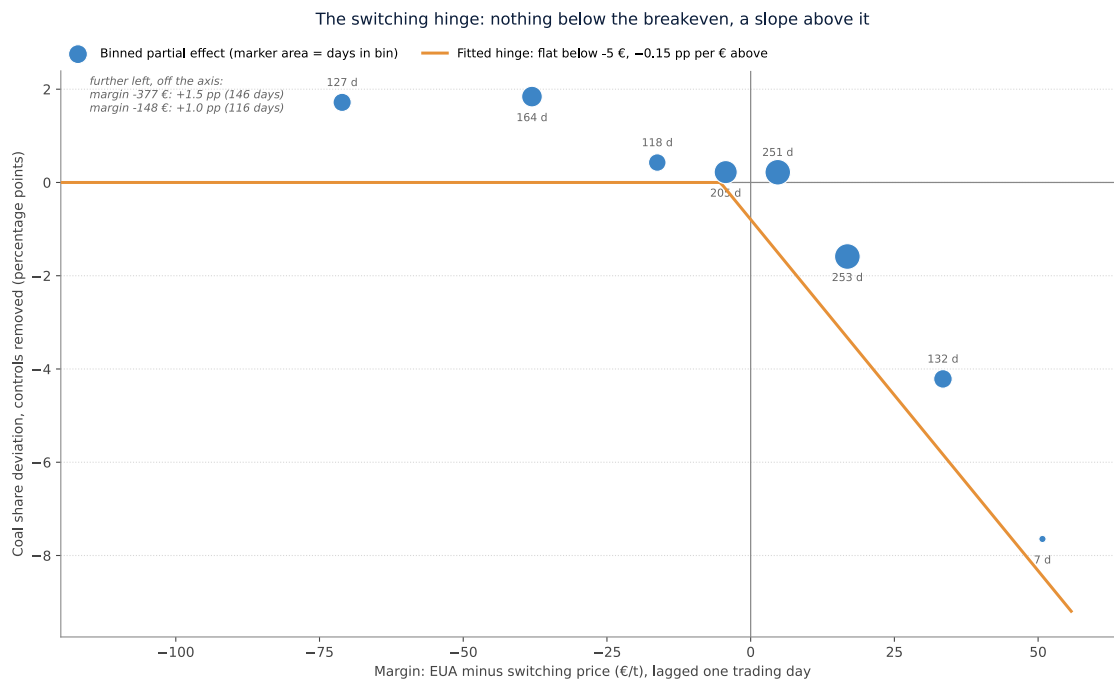


Figure 8 — Top: the allowance price and the coal-to-gas switching price, with the days the allowance sat above it shaded (42 per cent of 1,519 trading days; none in 2022). Bottom: the coal share of EU27 coal-plus-gas generation, 20-day mean. Switching price from the platform's fuel and efficiency assumptions; generation from ENTSO-E.

Measured on those 1,519 deduplicated trading days from the end of 2018 to September 2026, with the coal share of EU27 coal-plus-gas generation as the outcome and the margin of the allowance price over the switching price, lagged one day for day-ahead dispatch, as the driver, the response is a hinge (Figure 9). Below the breakeven the coal-fired generators run what they can and the share does not respond to how far below. Above it, every euro of margin is associated with about **0.15 percentage points** of coal-plus-gas generation moving from coal to gas — roughly **0.11 per cent of emissions from coal- and gas-fired generation per euro** — with the kink at **−5 €** (a small early-switching effect from the least efficient coal plants). Days with a positive margin are 42 per cent of the sample: none in 2022, nearly every day above the kink in 2020. The binned partial effects, controls removed, run from +1.8 percentage points of coal share at margins between **−50 and −25 €**, through zero at the kink, to

–4.2 at 25–50 € and –7.6 above 50 €. The design — a one-day lag, fuel and load controls, a threshold that is crossed and re-crossed, and a monthly re-run that cannot see it — is what makes the dispatch reading the natural one; the estimate itself is an association on observed days.



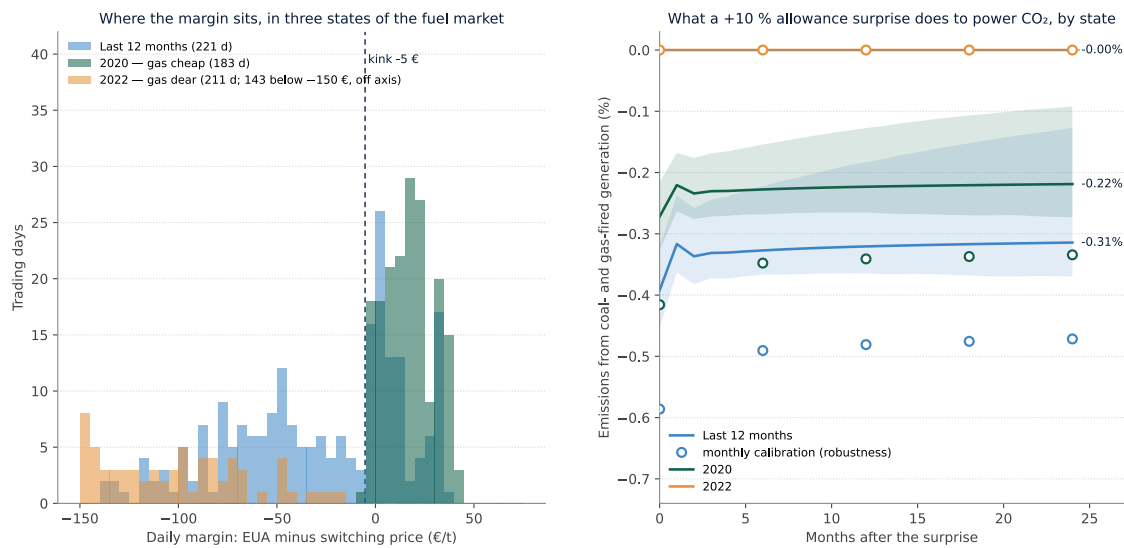
Historical — not a forecast, and not the model's horizon. 1,519 trading days, 2018-12-31 to 2026-09-04. Controls: log fossil generation, trailing-365-day share, month and weekday. Each euro above the kink moves ≈ 0.15 pp of coal-plus-gas generation to gas, $\approx 0.11\%$ of emissions from coal- and gas-fired generation. The hinge beats a linear fit on AIC.

switching/results.json 2026-09-07 · render_wp05.py E11

Figure 9 — The coal share's deviation from what the controls predict, binned by the day's margin (marker area is the number of days in the bin), with the fitted hinge: flat below the kink at –5 €, then 0.15 percentage points of coal-plus-gas generation moving to gas for every euro of margin above it. The two bins far to the left, at margins of –148 and –377 €, sit at +1.0 and +1.5 points: below the breakeven the distance does not matter.

Composed with the long-sample price response of \$4.3 (Figure 10): a ten-per-cent policy surprise on the allowance, about €8 at today's price, cuts coal-plus-gas power CO₂ by 0.3 per cent on the daily calibration and 0.5 on the monthly one — roughly 1.4–2.1 Mt a year — and holds there from month 6 to month 24 because the price effect persists and dispatch follows the level. The same surprise in a 2022-like state, with gas so dear that no coal plant sat at the margin, does nothing. This is the first result of the programme that composes two of our own measured pieces rather than replicating someone else's, and it carries the design's central lesson for the reader of scenarios: **the emissions consequence of a carbon price is state-dependent, and the state is set by the fuel markets, not by the carbon market.**

Price to emissions: the same surprise cuts emissions in one state and nothing in another



Historical — not a forecast, and not the model's horizon. Left: the daily margin in each state; the hinge acts only to the right of the kink. Right: the long-sample price response (a +10 % surprise, 8.2 % still in the price at 12 months, 8.0 % at 24) carried through the hinge over each state's own margins, with joint bands from the price and hinge uncertainty; lines use the daily hinge (0.3 per cent in the current state), the hollow circles the monthly-calibrated hinge (factor 1.53), the robustness estimate the text quotes alongside it. In 2022 no day was near the kink, so the response is nil whatever the price does; in 2020 nearly every day was above it but the allowance was €25, so ten per cent was €2.5 of margin against €8 today.

projection/results.json 2026-09-09 · switching/data/daily.csv 2026-09-04 · render_wp05.py E23+E24

Figure 10 — Left: where the daily margin sat in three states of the fuel market — the last twelve months, 2020 and 2022 — against the kink. Right: the long-sample price response to a ten-per-cent surprise carried through the hinge over each state's own margins, month by month, with joint bands from the price and hinge uncertainty; the lines use the daily hinge (0.3 per cent in the current state), the hollow circles the monthly-calibrated hinge (0.5 per cent), the robustness estimate the text quotes alongside it. In 2022 no day was near the kink, so the response is nil; in 2020 nearly every day was above it, but the allowance was €25, so ten per cent was €2.5 of margin against €8 today.

A monthly model cannot see this. Re-run at monthly frequency on the same period, the switching effect is not detectable: the threshold is crossed and re-crossed within the month and the monthly average sits near the kink (Figure 11). This is why the generator carries a daily dispatch layer, and how the daily result enters the monthly model is simple: on every simulated path the generator draws daily allowance and fuel prices, computes the day's switching margin, applies the hinge of Figure 9 day by day, and hands the month's *switching depth* — the average over the month's trading days of the margin above the kink, zero on days below it — up to the monthly emissions equation as a quantity with the hinge slope as its coefficient. Figure 11 is that quantity in history. The monthly equation is estimated on emissions with the switching effect taken out, and the simulation puts it back path by path, so one set of dispatch constants serves both places. The hinge is measured where it can be seen, at daily frequency, and used where it is needed, at monthly frequency.

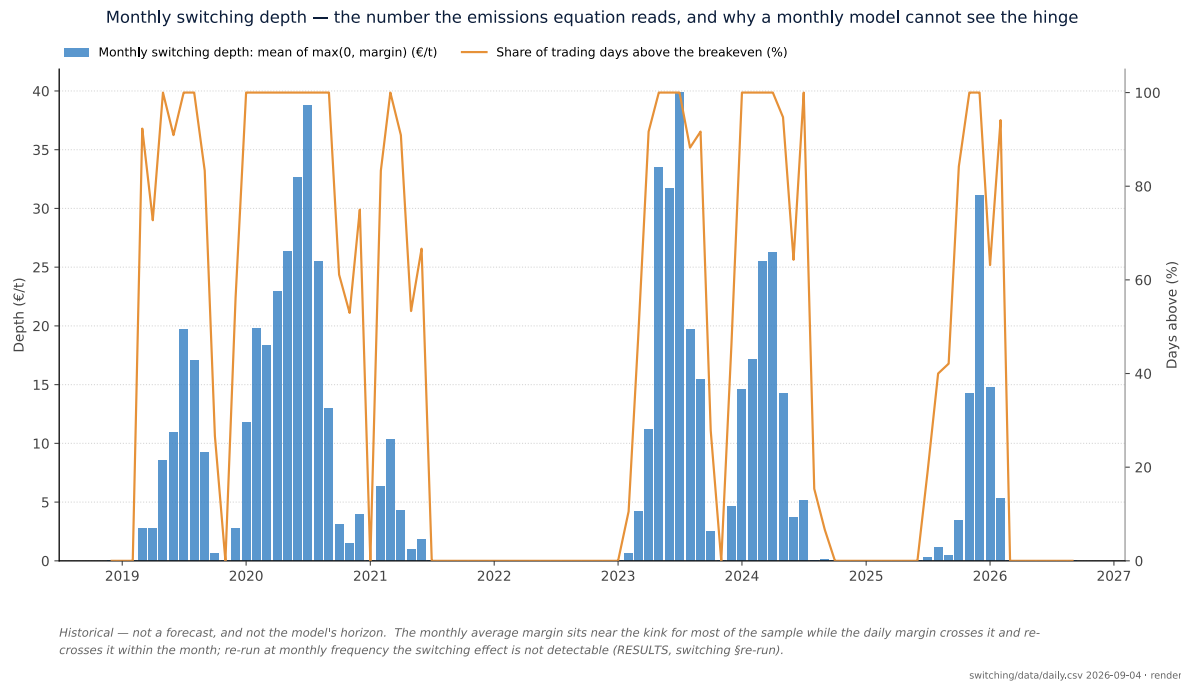


Figure 11 — The monthly switching depth (bars): the average over the month's trading days of the allowance's margin above the switching price, zero on days below it — the quantity the emissions equation reads. The line is the share of the month's days above the breakeven. Estimated at monthly frequency on the month's average margin, the hinge is invisible; computed from the days and handed up, it is not.

6. THE GENERATOR: LAYER BY LAYER, EACH TESTED ON THE BACKTEST

6.1 The monthly engine, and the first lesson

The engine is a three-variable monthly system — activity, power emissions, the real allowance price, with Brent, gas and coal at lags zero and one and heating and cooling degree-day anomalies — estimated on 2019 to the origin month, with the innovations drawn from its own residuals. The backtest re-estimates it at each of 52 month-ends from February 2022 to May 2026 on data to that month only, runs it forward, and scores the realised price at one, three, six, twelve and twenty-four months against a driftless random walk with the trailing two-year volatility.

The unrestricted system is a disaster (Figure 12). Estimated on a sample in which the price quadrupled while gas and coal rose with it, it learned that rise as its normal and projected it: explosive at 19 of its 53 origins (the levels engine has one origin more than the differenced one, 53 from January 2022 against 52 from February, because differencing gives up a month), losing to the random walk by a factor of 1.2 at one month, 4 at twelve and 16 at twenty-four, with the realised price below the fan month after month through 2023 and 2024. Shrunk toward a random walk with a Minnesota prior it recovers; moved into first differences with the drift shrunk hard toward zero it recovers fully — to near-parity with the random walk at every horizon (score ratios 0.99–1.03) and 80 per cent coverage of 0.83 at one month and 0.95–0.97 at six to twenty-four. Feeding in the realised fuel paths, as if they had been known, does not help at twelve months either: the fuel multipliers are real within the month but small against the price's own noise over a year.

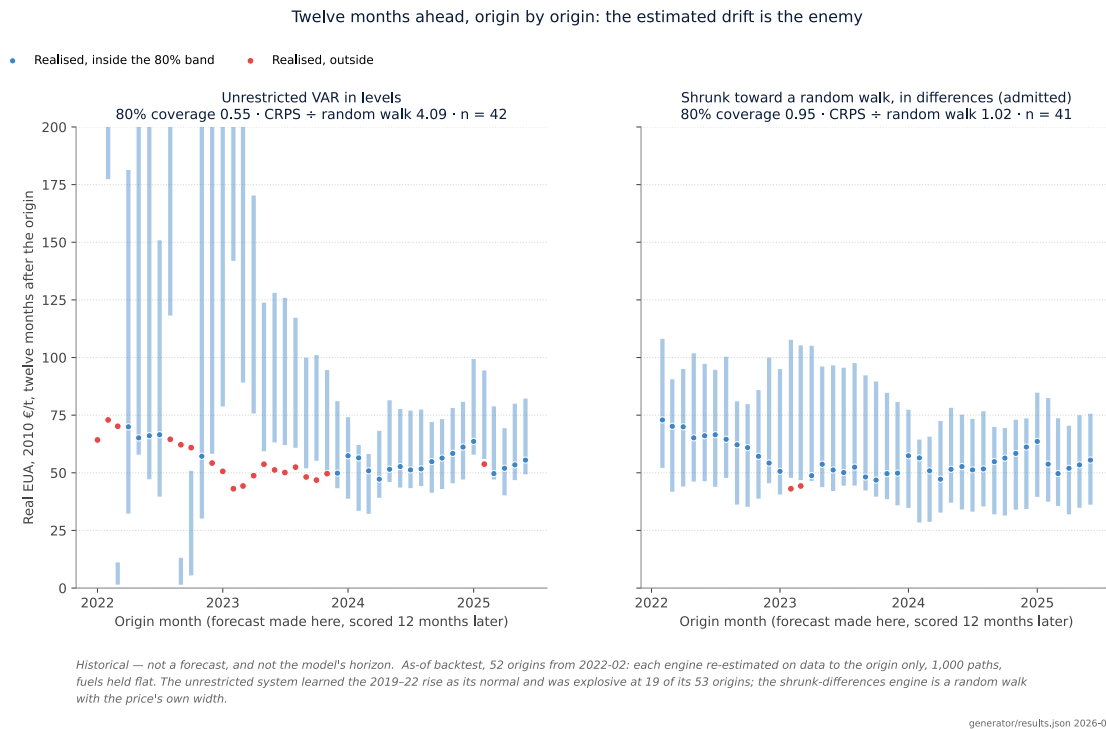


Figure 12 — The same test on the two engines. Each vertical bar is the 80 per cent band the engine gave, at that origin month, for the real price twelve months later; the dot is what the price then did (blue inside the band, red outside). Left, the unrestricted system in levels: through 2022 and 2023 its bands run off the top of the chart — it is projecting the 2019–22 rise forward — and from the December 2022 origin the realised price sits below them twelve months in a row; coverage 0.55, score four times worse than a random walk. Right, the admitted engine: bands of the price's own width, drifting with the level, score at near-parity with the random walk (1.02). Its coverage of 0.95 for a nominal 80 per cent band is not "the right width": it is a stable, benchmark-like width that over-covered in this particular sample, as 7 records. The engine earns no forecasting skill on the level.

What the shrinkage does deserves a plain statement, since the word carries the argument. Each of the three equations has twelve coefficients and, at the early origins, only three or four years of monthly data to estimate them from; ordinary least squares fits whatever the sample happens to contain, including the accident that price, gas and coal all rose together through 2019–22. A Minnesota prior — named for the Federal Reserve Bank of Minneapolis, where Litterman introduced it in the 1980s — is a rule for how far each coefficient may stray from a stated benchmark, here a random walk: in a system estimated in differences, every coefficient's benchmark is zero, and the estimate is a compromise between the data and that benchmark governed by one tightness parameter, λ . At $\lambda = 0$ the engine is a random walk; as λ grows the estimate returns to least squares; in between a coefficient moves away from zero only in proportion to how strongly the data insist, each judged against the relative volatility of the series involved. The engine uses $\lambda = 0.2$, one lag, and a much tighter zero-centred prior on the intercept — the drift — because a drift estimated on a sample that quadrupled is small in any month and compounds over twenty-four, and the backtest showed nothing short of that cures it; the formulas are in the methods note. The honest reading of the right-hand panel of Figure 12 is that shrinkage buys near-parity with a random walk, not skill; what the engine contributes is a stable, benchmark-like width and the propagation of shocks between the variables, not knowledge of where the level is going.

The lesson is stated once and applied everywhere: **an estimated drift is the enemy**. The monthly engine does not know where the price level should be; the level is the market's, the engine's job is propagation and width. Three production rules follow: the engine runs in differences under a

random-walk prior with the drift shrunk hard toward zero; residuals are centred before they are drawn (an uncentred bootstrap put the shrunk-away drift straight back in); and no engine is simulated with a largest root above 0.995.

6.2 The daily layer, and what it is for

The design hands the price's within-month randomness to the daily clock: ordinary noise plus policy jumps, with the monthly engine supplying the month-to-month innovation. Built as specified: jumps arrive at the rate the register shows since 2015 — 118 eligible events in 3,010 trading days, about ten a year — with sizes drawn from those events' own surprises (standard deviation 2.5 per cent, largest 8) and full persistence — the estimate of \$4.3 is below one and imprecise, and full persistence is taken as the conservative production assumption, since a jump that fades would narrow the fan; ordinary noise is a block bootstrap of non-event daily returns, scaled once so that its monthly aggregate matches the engine's innovation (scale 0.93; the jumps carry 4 per cent of the monthly variance). The process reproduces the daily record's fat tails: kurtosis 4.2 against 4.2, one day in twenty moving more than five per cent against one in sixteen (Figure 13).

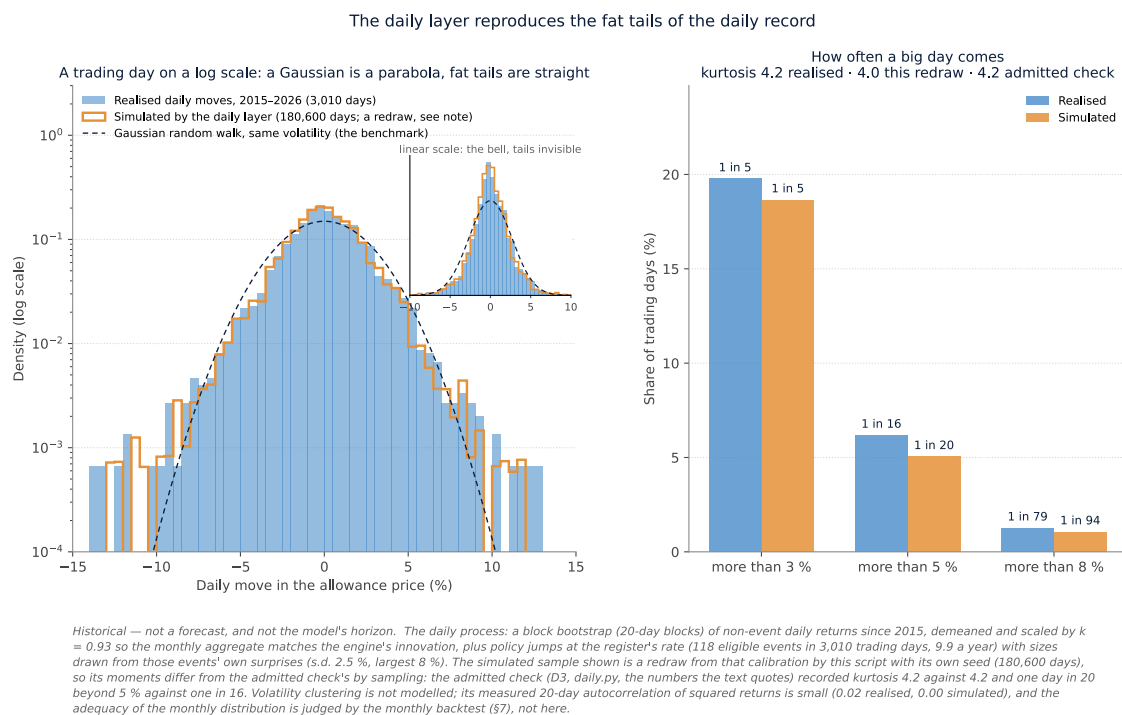
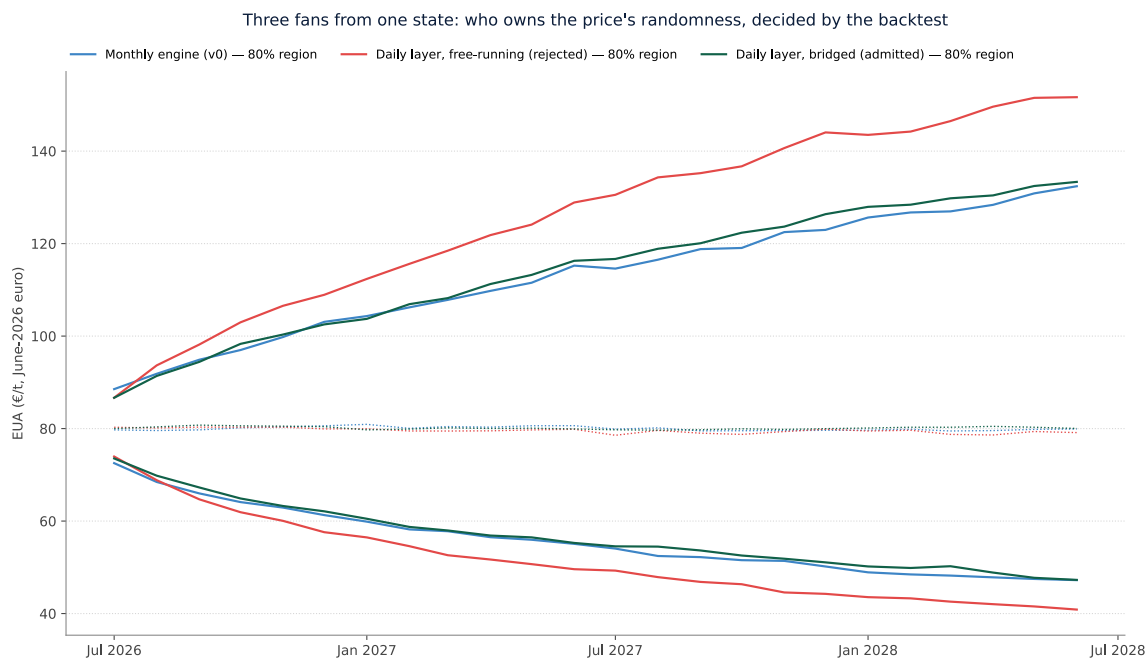


Figure 13 — Left: the distribution of daily moves in the allowance price since 2015 (blue, 3,010 days) against what the daily layer produces (orange outline; a redraw from the same calibration, so its moments differ from the admitted check's by sampling — both are in the panel title), on a log scale so the tails can be seen; the dashed curve is a Gaussian of the same volatility — the increments of the random-walk benchmark used in the backtest. On a log scale the bell becomes a parabola (the inset shows the ordinary view), and fat tails show as a straight fall-off where the parabola keeps bending down: a move beyond 10 per cent, which the Gaussian puts at once in twenty-two years, has happened sixteen times in twelve. Right, how often a big day comes: the simulation matches the record beyond 3 per cent and slightly under-produces it beyond 5 and 8. Volatility clustering is not modelled; its measured 20-day autocorrelation of squared returns is small (0.02), and the adequacy of the monthly distribution is judged by the backtest of \$7, not here.

Then the backtest changed the ownership rule (Figure 14). Run free — the daily process carrying on across months — the fan is wider than the monthly engine's by a quarter at six months, covers 100 per cent, and loses to the random walk by 12–16 per cent, for a sixty-year-old reason (Working, 1960):

monthly averages of a daily random walk have positively autocorrelated innovations, so their fan grows faster than the square root of time, and the realised monthly series does not behave that way. Run as a *bridge* — the daily layer owns the current month, where it knows the last close rather than the month's average, and the within-month path thereafter, while the engine's own draw sets each later month's average — the fan is the engine's from six months out. The bridge's apparent one-month advantage in the layer test was information, not skill: with the benchmark given the same last close (\$7) it is at parity at one month, as at every other lead, and it is retained as the correct treatment of what is known at the moment of the forecast, not as a source of forecasting advantage.



Rolling two-year trading-horizon model — not a strategic projection; the reader's exposure is usually longer than the fan's reach. Backtest, 52 origins from 2022 — monthly: CRPS+RW 1.00 at 1 m, 1.03 at 12 m; 80% cov 0.95 free: CRPS+RW 0.74 at 1 m, 1.16 at 12 m; 80% cov 1.00 bridge: CRPS+RW 0.74 at 1 m, 1.03 at 12 m; 80% cov 0.95. The free-running daily process makes monthly averages with positively autocorrelated innovations (Working, 1960) and over-covers; the bridge owns the first month and the within-month path and is otherwise the engine. The one-month 0.74 is against a random walk that does not know the last close; given the same close the bridge is at parity (\$7).

generator/daily_results.json 2026-09-19 · render_wp05.py E26

Figure 14 — The 80 per cent region of the price fan from the June 2026 state under the three ownership rules, with the medians dotted. Blue, the monthly engine alone. Red, the daily process left to run across months: the same process, aggregated to monthly averages, produces a fan a third wider at twelve months than the engine's and, in the backtest, covered everything and scored 12–16 per cent worse than a random walk. Green, the bridge: indistinguishable from the engine from six months out, as it must be, since from the second month the engine sets every month's average. The backtest columns are in the note; the bridge's one-month score is against a benchmark that does not know the last close, and at parity against one that does.

How the bridge works, in one pass: the monthly engine deals in monthly averages and never sees a closing price; the daily layer deals in closes. In the first month the daily layer is in charge — it starts from the last close actually observed, which the engine does not know, and the average of the month it simulates becomes the engine's price innovation for that month. From the second month on the engine fixes each month's average from its own residuals, and the daily layer, continuing from wherever its path ended, supplies the shape inside the month, tilted so that its average lands on the engine's figure. The month-to-month movement is the engine's, the day-to-day tails are the daily layer's, and the two are consistent by construction; the tilt is what removes Working's problem.

Once the generator moved into daily production a further piece was needed and tested the same way: at any day inside a month, the realised month-to-date path is taken as given and only the remaining

trading days are simulated. Tested at 56 mid-month origins, the partial month's score is a third of a full month's, with coverage of 0.86.

6.3 Fuels, dispatch, and the emissions line

Fuels — what was tested. Brent, gas (TTF) and coal (API2) are drawn jointly — blocks of twenty trading days from their joint history since 2019, so that a gas spike arrives with the coal and oil moves that came with it — around a slowly mean-reverting level: each fuel is pulled toward its trailing three-year average at a speed estimated as-of, the pull phased in over the first month. Four alternatives were tested on the fuels themselves at 55 origins from December 2021. A calendar seasonal for gas was a quarter worse than the random walk at one month and a sixth worse at three, and was rejected. The driftless random walk gives gas a two-year 80 per cent region of €18–306 and was rejected as the unconditional planning baseline. Two variants driven by the fuels' own forward curves — the Brent, TTF and API2 curves, outright and strips, which the platform delivered with their full month-end history from 2019 — were tested: one in which each fuel drifts along its curve, one in which it reverts toward the curve instead of toward its trailing mean.

What the tests found, and the ruling. For gas, both curve variants were 4–9 per cent worse than the random walk at one to six months and 25–31 per cent better at twenty-four; both also covered 100 per cent at twenty-four months, because the curve's drift is informative while the random-walk noise around it stays at full width; and both therefore failed the pre-stated admission rule. The trailing-mean reversion missed the same rule by 0.0006 at twelve months, with well-behaved coverage and gains that persist at longer horizons; that miss was judged economically immaterial and the reversion adopted by an explicit owner ruling, whereas the curve variants failed more substantively. On this sample the curve's gain was earned in 2021–22, when the end-2021 curve priced the unwinding of the gas spike two years out and was right, while the same curve was twice as bad as a random walk at twelve months in 2021 and 28 per cent worse in 2023. Forward prices carry risk premia and hedging demand as well as expectations, so nothing here is offered as a general property of forward curves; whether the near-term bias shrinks as the market matures is a hypothesis the forward scores will test. The ruling: mean reversion as the unconditional fuel process; "fuels follow the forward curve" as a named scenario produced on every run beside it; and a written trigger for a formal model review — not an automatic switch — under which the TTF curve may become the gas-price anchor once at least twenty-four production forecasts have twelve-month outcomes, the curve scenario's mean twelve-month gas score is at least 5 per cent below the reversion's, and its coverage is no worse and within the accepted range. Those twenty-four overlapping origins hold about two independent years, so the review weighs the difference with the treatment of overlapping errors used in §7, adoption remains an owner ruling, and the first possible review falls around late 2029. Brent and coal remain mean-reverting unless separately tested and ruled upon.

Curves not used here. Two further curves the platform holds — the EUA December futures curve and the AAA euro-area government spot curve with the €STR series — do not enter the two-year generator at all. They are held for the separate long-horizon layer of the specification, where the EUA curve will measure the carry the market charges for holding carbon and the euro curve will separate the risk-free rate from that premium; that layer is not yet operating, because the bank ledger it rests on is incomplete.

Dispatch. In an electricity system, dispatch is the allocation of generation across the units available in each hour according to their running costs; for coal- and gas-fired units those costs are the fuel and the allowances each burns, so the allowance, gas and coal prices are central to determining how

much of coal-plus-gas generation comes from coal, alongside load, renewable output, plant availability and network constraints. The generator carries this as the aggregate, reduced-form coal-to-gas switching model of §5 — a hinge estimated on the system's daily coal share — not as a unit-level simulation of the power market, and no power station switches fuel: the shift is between coal-fired and gas-fired units. Each simulated day has a gas, a coal and an allowance price; the first two imply the switching price; when the allowance sits sufficiently above it the hinge shifts generation from coal toward gas by the measured amount per euro of margin; and the month's average margin — the switching depth of Figure 11 — enters the power-emissions equation. From the September 2026 state, with gas near €77 and the switching price near €217, the median path never crosses the threshold in two years, while the paths on which gas falls toward its three-year average do — in the spring of 2027 for the lowest quarter — and the tenth percentile of the emissions index at twenty-four months is 30 per cent below today's level, the ninetieth 8 per cent above (Figure 15). That one-sidedness is the emissions fan's downward skew.

What the fuel assumptions change. Making the fuels uncertain changes the allowance-price fan — the EUA fan, the paper's principal object — hardly at all: in the June 2026 test its 80 per cent region at twelve months moved by about €1–2 (€56–116 with fuels stochastic against €55–114 with them held), and its as-of backtest was unchanged, within 0.02 of the score ratio at every horizon. The reason is in the engine's coefficients: the estimated effect of a fuel move on the allowance price within the month is real but small, and over a year it is swamped by the allowance's own price innovation, which the random-walk prior leaves at full size. What the fuel assumptions do change, materially, is the power-emissions fan, because the fuels enter the emissions equation directly and set where the switching threshold sits relative to the allowance price; between the fuel variants tested, the dominant difference in the emissions fans is the gas path.

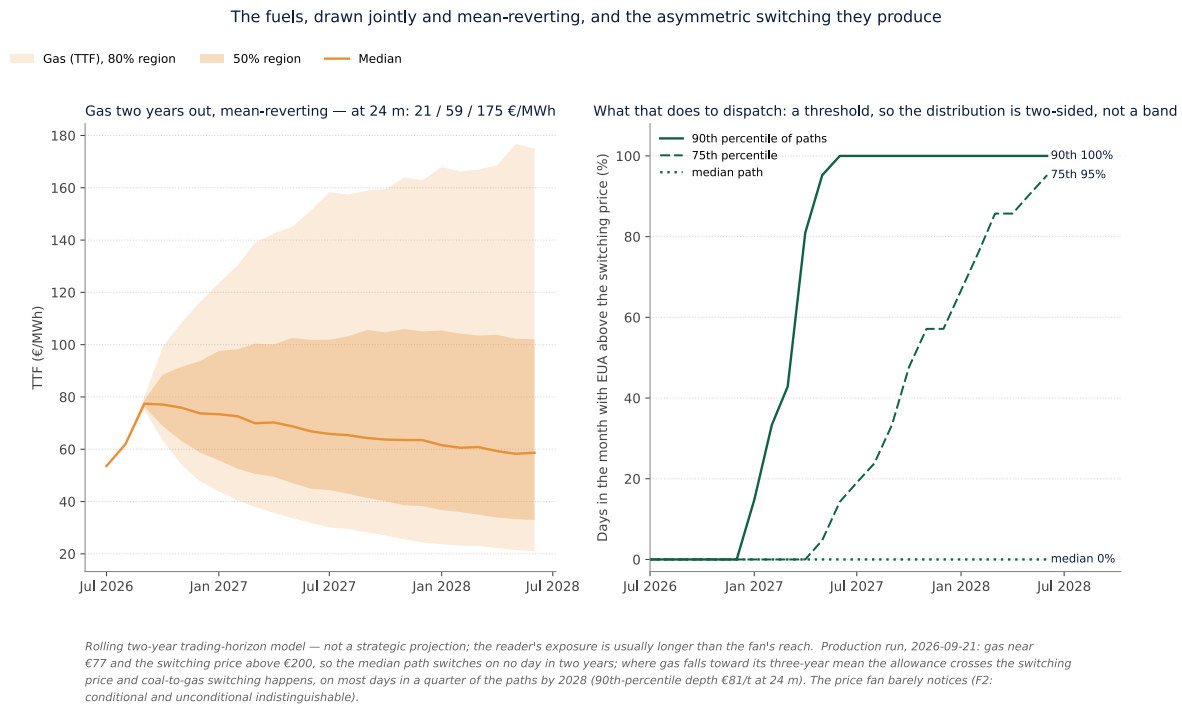


Figure 15 — Left, gas two years out from the 21 September 2026 run: the median drifts from €77 to €59 by mid-2028, pulled toward a trailing three-year average near €40 at the speed the data allow; the 80 per cent region reaches €21–175 by mid-2028 — the width is gas's own history since 2019, the 2022 spike included. Right, the consequence for coal-to-gas switching, read as the share of trading days in each month on which the allowance price is above the switching price implied by that day's gas and coal prices. It is a threshold, not a band around a central path: the median path never crosses it in two years (0 per cent of days, the dotted line along the floor), while in the quarter of paths where gas falls furthest the crossing arrives in the spring of 2027 and switching is happening on nearly every day by 2028. This one-sidedness is why the power-emissions fan of \$6.4 is skewed downward while the allowance-price fan is not.

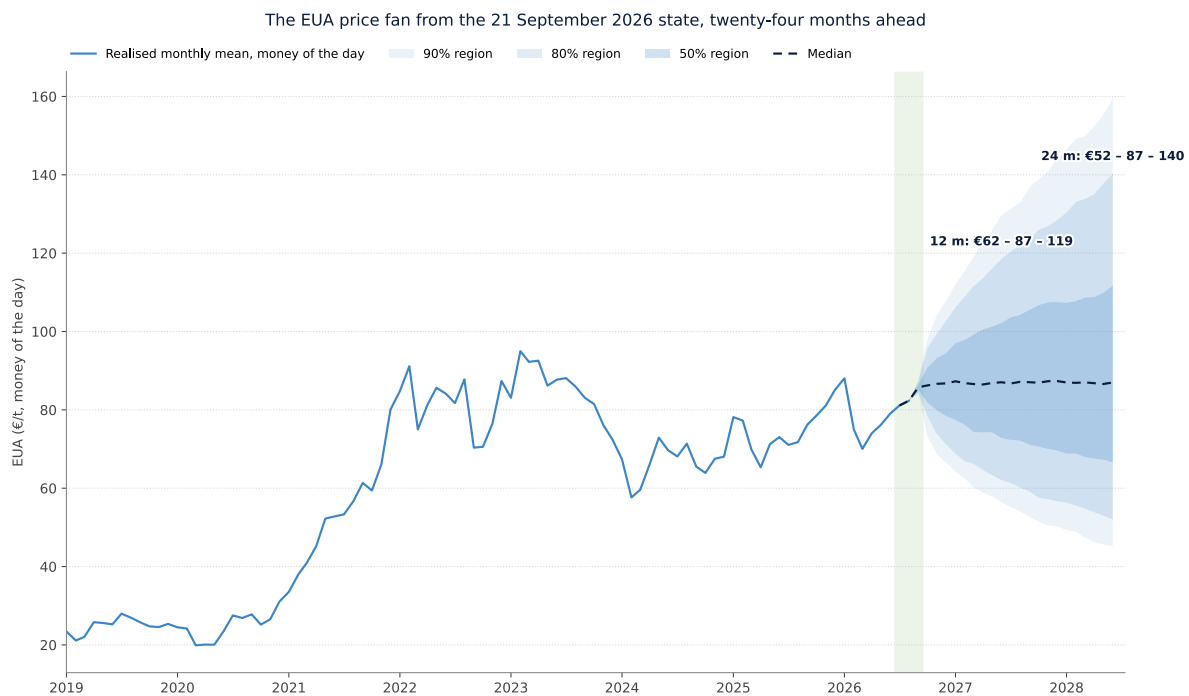
Emissions. Power-sector emissions inherit nothing from the price's random-walk prior; they have the equation their data ask for — a level that reverts around a trend (last month's level carries a coefficient of 0.44; the long-run trend about –7 per cent a year), with activity, the price, the fuels and weather as regressors, re-estimated at every origin and re-fitted on every simulated path so that the estimation uncertainty is in the fan. Tested as-of against a random walk and a one-variable trend-with-reversion it earned its place in the layer test: 24 and 38 per cent closer to the truth than the random walk at twelve and twenty-four months. Left to its in-sample residuals it was over-confident (80 per cent bands catching the outcome 61–68 per cent of the time); the production version scales the innovation by a recursive calibration factor — the root-mean-square of the last two years' one-step errors, each measured against the equation as it stood the month before — which lifts coverage to 0.71–0.83 without moving accuracy. Shrinkage was retried on this equation with the textbook scaling and rejected at every strength: shrink where nothing beats a random walk, do not shrink where the regressors have content. Activity itself is a pure random walk with two-year volatility, since nothing beat that. What the consolidated test says about *when* this equation had skill is in §7.

6.4 The fan from the September 2026 state

Run by the production package on the data as of 21 September 2026 — June 2026 model origin, the last complete panel month, updated through 21 September with July and August realised and September month-to-date — 2,000 paths (Figure 16; the fuels and the switching they produce, Figure 15; the emissions fan, Figure 17):

horizon	EUA, money of the day, 10th – 50th – 90th percentile	power CO ₂ vs today	gas (TTF) €/MWh
3 months (Dec 2026)	€84.5 – €85.8 – €87.2	–13 % – +4 % – +21 %	75 – 77 – 80
6 months	€71 – €87 – €103	–14 – +2 – +21	48 – 74 – 116
12 months (Jun 2027)	€62 – €87 – €119	–21 – –3 – +17	32 – 67 – 151
24 months (Jun 2028)	€52 – €87 – €140	–30 – –13 – +8	21 – 59 – 175

The centre is flat because the model does not know where the level should be (\$6.1) and says so; the width is the price's own history, 2022 included, and it is the model's central 80 per cent interval — a deliberately conservative envelope, by ruling, not a claim that the price will fall inside it eight times in ten (\$7 reports what it did). The three-month numbers are narrow because the months to December are largely realised. The emissions fan's downward drift is the trend and the state of the gas market; its upper tail is little or no coal-to-gas switching, its lower tail coal-to-gas switching on most days.



Rolling two-year trading-horizon model — not a strategic projection; the reader's exposure is usually longer than the fan's reach. Origin 2026-06 (last complete panel month); 3 months realised or month-to-date (shaded); 2,000 paths; unconditional mode (fuels mean-reverting, events at the calibrated rate); layers not yet built: weather_bootstrap, pending_dossiers, bank_identity. The centre is flat by construction (\$6.1); the width is the price's own history, 2022 included.

2026-09-21b · p202609-20260921-0.4.1 · svar_generator 0.4.1 · render_wp05.py E18

Figure 16 — The price fan itself, in money of the day, against the realised monthly mean since 2019. The shaded strip marks the three months between the June model origin and the run date, which are realised or month-to-date and therefore nearly certain; the fan proper begins in October 2026. The median runs flat at about €87; the 80 per cent region reaches €62–119 at twelve months and €52–140 at twenty-four, where the 90 per cent region is €45–160. Set against the history to its left the width is not extravagant: the monthly mean went from €34 in January 2021 to €95 in February 2023 and back to €58 a year later.

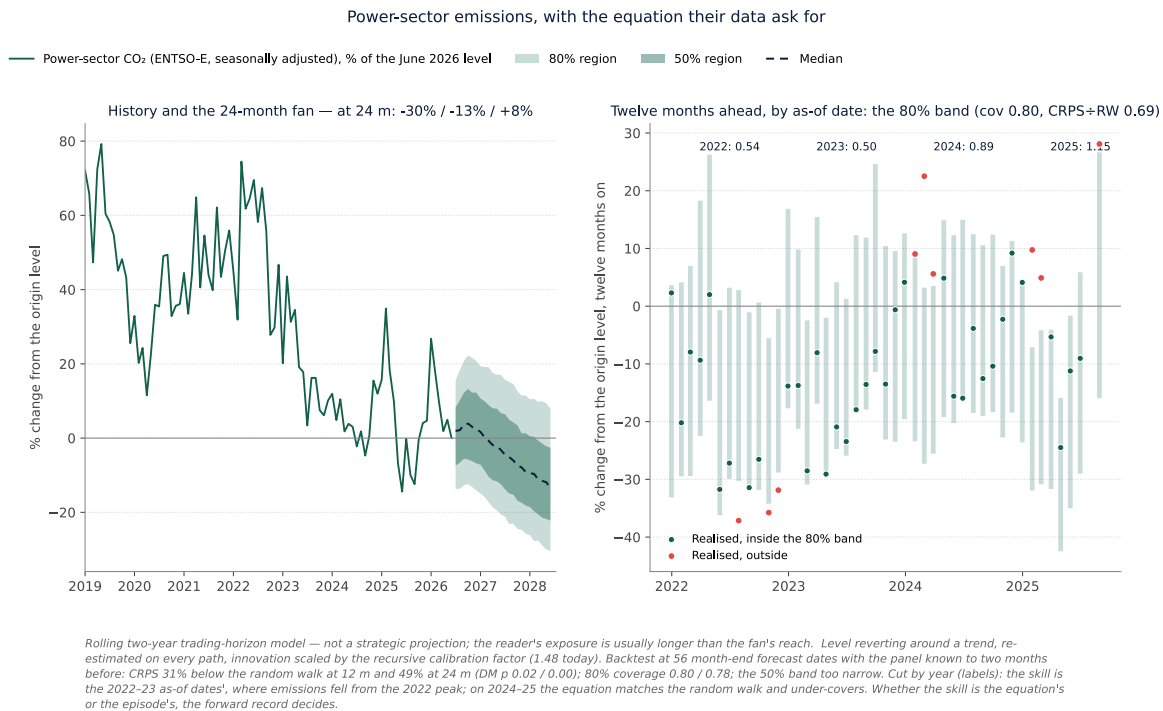


Figure 17 — Left: power-sector CO₂ since 2019 as a percentage of the June 2026 level, and the production fan to mid-2028 — a median falling 13 per cent by two years, the 80 per cent region -30 to +8, skewed downward for the reason §6.3 gives. Right: the same equation's twelve-month forecasts at each forecast date of the backtest of §7, the 80 per cent band as a bar in the legend's lighter shade and the realised outcome as a dot, red where it fell outside; the labels are the error ratio to the random walk for each year's forecast dates. The bands are wide and the outcomes inside them through 2022-23, when the equation was ahead of the random walk; from 2024 the misses cluster and the ratio rises to one and above.

7. TESTING THE FORECAST: WHAT WORKED AND WHAT DID NOT

Section 6 described how the generator was assembled and how each component was tested before inclusion. This section tests the resulting production model as a whole. The backtest recreates the information available at each of 56 month-ends from December 2021 to July 2026 and compares the resulting forecasts with a driftless random walk given the same information.

An earlier version of the backtest overstated the model's near-term performance because the model was given the final observation for the month while the benchmark was not. That result has been withdrawn. The results below use the corrected equal-information design.

At each forecast date the model knows daily prices, fuels and events to that day's close, and the monthly panel only to two months before: industrial production is published about 45 days after month-end and power emissions two to three months after, so the two months in between are realised for prices and fuels and simulated for activity and emissions. Archived historical vintages of the seasonally adjusted panel do not exist, so this is a lagged latest-vintage backtest rather than a true real-time one, and it is labelled so. The benchmark is a driftless random walk started from the same close, given the same information (its construction is in the methods note).

Forecast quality is assessed in three ways: whether the stated intervals contained the outcome at approximately the promised frequency; whether outcomes were systematically concentrated in one

part of the distribution; and whether the complete distribution was more accurate than the random walk, measured by the continuous ranked probability score and read as a ratio to the benchmark's. Differences in accuracy were tested allowing for the overlap between successive forecasts. Full definitions, test specifications and the complete results table are in the methods note, available on request.

forecast	horizon	error relative to random walk	80 % coverage	interpretation
EUA price	12 / 24 months	1.02 / 1.01	0.96 / 1.00	no forecasting advantage; wide in this sample
power CO ₂	12 / 24 months	0.69 / 0.51	0.80 / 0.78	strong pooled result, concentrated in 2022–23
gas (TTF)	12 / 24 months	0.96 / 0.94	0.84 / 0.88	modest advantage, not statistically established



Historical — not a forecast, and not the model's horizon. As-of = the month-end close; the panel known to two months before; benchmark = a close-anchored random walk with the monthly variance for whole months and the daily variance within the target month. Flat PIT histogram = calibrated; piled in the middle = too wide; leaning = biased. Price: parity at every lead (1.01–1.02), and the benchmark's own 80% band covers 0.96 / 1.00 at 12 / 24 m, as the fan's does — the over-coverage is the sample's; emissions: beats the random walk from 3 months, on the 2022–23 as-of dates (E28); gas: 0.94–1.01, too high in 2023–25. Seeds 2 and 3 agree to ± 0.01 .

generator/g4c_results.json 2026-09-25 (seed 1; s2, s3 stability) · render_wp05.py E30

Figure 18 — Top: the error ratio to the random walk by lead for the three variables; a ratio of one is the dashed line, and a filled marker means the difference in accuracy is significant at 5 per cent once the overlap between forecasts is allowed for. The price sits on the line at every lead; gas drifts a little below it; power emissions fall from 0.85 at one month to 0.51 at two years, significantly from three months on. Bottom: where the outcome fell in the fan, at twelve and twenty-four months, against the flat line a calibrated fan would produce. The price piles in the middle — the band was wider than the outcomes required; emissions are close to flat; gas leans left because its fan sat above the outcomes in 2023–25.

The EUA price forecast did no better than the random walk, and both were wider than the period required. The error ratio is 1.01–1.02 from one month to two years and within a few hundredths of one in every year of forecast dates (Figure 19); the fan's centre is where the random walk's is. A ratio of one does not mean the model adds nothing — it carries the emissions line, the scenario structure

and the tails — but it does mean the model showed no improvement in predictive accuracy over the benchmark in this sample — what one would expect if the price were close to a martingale at these horizons, a hypothesis this sample cannot distinguish from small predictable components. On width, the acceptance criterion written before the test allowed the share of forecast dates on which the outcome fell inside the 80 per cent band to lie between 0.65 and 0.95; the price band scored 0.96 at twelve months (43 of 45) and 1.00 at twenty-four (all 33). **On that criterion the band fails at both leads: it was too wide for this period.** Two things qualify the failure without erasing it. Successive twelve-month forecasts overlap by eleven months, so 45 scores carry the information of about four independent windows, on which the excess is not statistically established. And the same over-coverage appears in the random-walk benchmark, suggesting that it reflects the unusually calm evaluation period rather than a defect unique to this model. What was done about it is in §7.1.

The emissions forecast was materially better than the random walk overall, but the advantage is concentrated in forecasts made during the 2022–23 decline. Pooled, it is 31 and 49 per cent closer to the truth at one and two years, with 80 per cent coverage no test distinguishes from nominal, though its 50 per cent band is too narrow. Cut by the year of the forecast date (Figure 19), the twelve-month error ratio is 0.54 for the 2022 dates, 0.50 for 2023, 0.89 for 2024 and 1.15 for the seven 2025 dates; at twenty-four months 0.40, 0.43 and 1.06. An equation that reverts a level around a falling trend was positioned to see the fall from the 2022 peak; in the calm 2024–25 dates it does no better than the random walk and under-covers. On the twelve non-overlapping twelve-month subsets the ratio runs from 0.45 to 1.18, ten of twelve below one, and every twenty-four-month subset that beats the random walk starts in 2022–23. Under the pre-specified acceptance criterion the emissions component passed on the pooled sample; the sentence the reader should carry is the qualified one: **the observed skill is concentrated in forecasts made during the decline from the 2022 peak, and none has yet been shown in a calm year.** Whether it is the equation's or the episode's is what the forward record will settle.

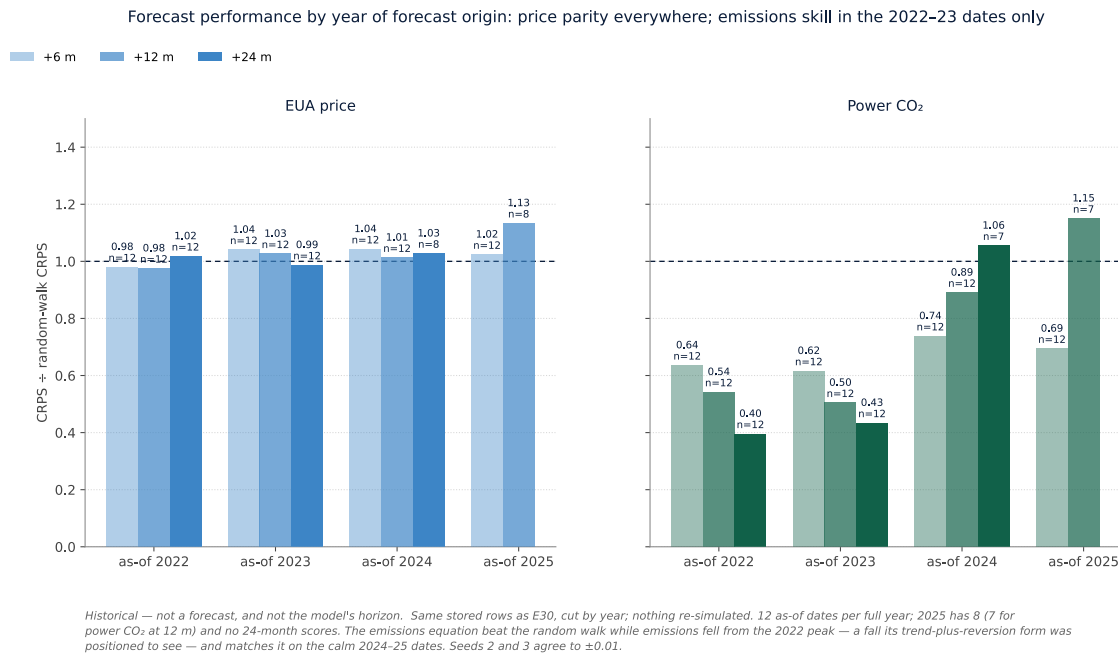


Figure 19 — The same stored scores as Figure 18, cut by the year of the forecast date and nothing re-simulated: the error ratio at six, twelve and twenty-four months for the price (left) and power emissions (right). The price is within a few hundredths of one in every year and at every lead. Emissions beat the random walk by half in the 2022 and 2023 dates, by a tenth at twelve months in 2024, and not at all in 2025 (seven dates, no twenty-four-month scores yet): the observed skill is concentrated in forecasts made during the decline from the 2022 peak.

Gas is reported for completeness: 0.94–0.96 against the random walk at twelve and twenty-four months, a modest advantage that is not statistically established, and a fan that sat too high in 2023–25 because its long-run mean was learned on a sample containing 2022.

7.1 Could the price fan be narrowed?

A data-driven adjustment was tested to see whether the price fan could be narrowed without damaging forecast accuracy: a scale factor on the engine's innovation, computed at each forecast date from the engine's own recent one-step errors, under an admission rule recorded before it was run. Run on the corrected backtest — the same fifty-six forecast dates, the same information set and the same random draws as the results above — it narrowed the band by eight per cent at twelve months and by two per cent at twenty-four, and left the coverage where it was: 0.98 and 1.00 against 0.96 and 1.00 without it. It was not admitted. The mismatch was too large for any such adjustment to solve: over the evaluation period the fan was roughly three times as wide as the errors its median encountered. The construction, the admission rule and the figure are in the methods note.

The alternative would be to fit mean reversion in the price to the relatively calm period since 2023. That was rejected because the sample contains too few independent long-horizon observations — about four at twelve months and about one at twenty-four — to establish that the market has entered a permanently quieter regime, and because another 2022-style shock is precisely what the fan exists to cover. The production fan therefore retains the 2022 shock in its risk envelope; a change of regime is expressed as a named stress scenario beside the unconditional fan, never by narrowing it. This is a conservative risk-management decision, not a claim that the central 80 per cent interval is currently calibrated to contain exactly 80 per cent of outcomes.

7.2 The record from here

Each production forecast is stored with its data, code and parameter versions and will be scored as outcomes become available. Because successive twelve-month forecasts overlap, robust evidence at that horizon will accumulate over years rather than months; the question of whether the price fan is too wide or the period was calm will be answered by that record, not by this sample.

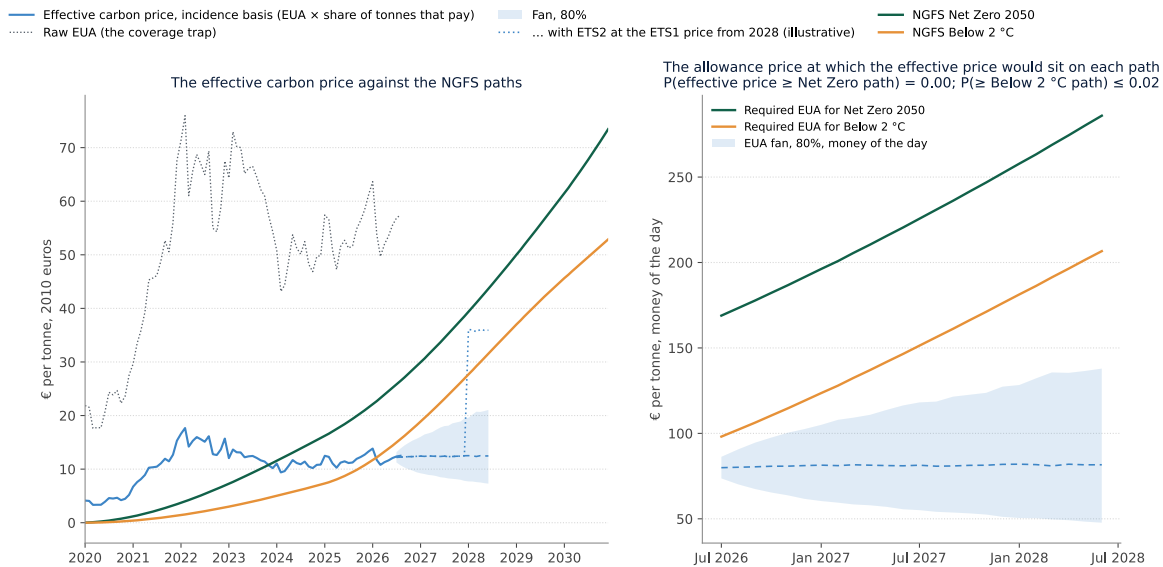
8. READING THE FAN AGAINST THE NGFS SCENARIOS

Insurers' ORSAs and banks' climate stress tests are commonly read against the NGFS scenarios — supervisors refer to them as a reference set rather than prescribing them — and every such scenario carries a carbon price. It is tempting to compare that price with the allowance price directly. The comparison is wrong by a factor that has nothing to do with markets.

The NGFS carbon price is an economy-wide shadow price from an integrated assessment model, applied to every tonne the economy emits. The allowance covers about 38 per cent of EU emissions, and once free allocation is netted out the tonnes that actually pay are 22 per cent. Compared raw, the market looks far above Net Zero — a "premium" that is almost entirely the logarithm of one over the coverage share. The comparable object is the **effective carbon price**: the allowance price weighted by the share of the economy's emissions it reaches, on two bases that answer different questions. The **incidence** basis — tonnes that pay, 22 per cent in 2024 — is the cash-cost measure, the right one for a provision, a cost line or a CBAM liability, and the primary basis here. The **marginal** basis — tonnes covered, 38 per cent — is the coverage-adjusted price signal: free allocation does not remove the opportunity cost of a tonne, so for investment incentives or abatement economics it is the one to read. Both are in 2010 euros, as the NGFS paths are. NGFS is the benchmark the fan is measured against, never an input to the model. One mismatch shapes how the whole comparison should be read: the scenario path used is the GCAM 6.0 EU-15 region of NGFS Phase 5, while the coverage shares are EU27. No EU27 path for the same model was available at the time of writing, so what follows is an illustrative scenario comparison — a measured effective price against one model's carbon-price path for a neighbouring region — not an exact score of the EU's alignment; an EU27 sensitivity will be added when the platform delivers the path.

The scorecard reads plainly (Figure 20), with that caveat attached to every number in it. On the incidence basis the probability that the EU's coverage-adjusted carbon price reaches or exceeds the Net Zero 2050 carbon-price path — which is what the scorecard computes; it is not the probability of being on a climate outcome — is below 0.2 per cent at every horizon, and against Below 2 °C it is between 0.3 and 2 per cent: the effective price today is €12.5 against a Net Zero path at €25.8 in 2010 money, rising to €43.6 by mid-2028, and the fan does not reach it. For the two to coincide the allowance would have to be **€169 today and €286 by mid-2028** in money of the day — a threshold-equivalence, not a required market price; the fan's 90th percentile at two years is €140. On the marginal basis the picture is closer — about even odds against Below 2 °C a year out — and on the raw comparison the market "beats" Net Zero with probability 0.75 to 1.0, which is the trap the table is there to show. The number to take away is the threshold-equivalent allowance price, because it makes the choice of basis transparent rather than hiding it inside a probability: the required allowance price depends directly on whether the exposure is measured on the 22 per cent incidence basis or the 38 per cent marginal one, and a reader who prefers the other basis can rescale it.

Reading the fan against the NGFS carbon-price paths — an illustrative comparison: the distance is coverage first, price second



Rolling two-year trading-horizon model — not a strategic projection; the reader's exposure is usually longer than the fan's reach. Shares: incidence 0.22, marginal 0.38 (2024, carried forward); NGFS Phase 5 v5.1, GCAM 6.0, EU-15. The ETS2 line is an upper bound with strong conditions (§8). Every probability is conditional on the scenario's non-policy assumptions holding.

generator/scorecard_results.json 2026-09-21 · render_wp05.py E29

Figure 20 — An illustrative comparison: a GCAM EU-15 path against EU27 coverage shares. Left: the effective carbon price on the incidence basis since 2020 (the allowance price times the share of the economy's tonnes that pay), its two-year fan, and the NGFS Net Zero 2050 and Below 2 °C carbon-price paths, all in 2010 euros; the dotted grey line is the raw allowance price, which sits above Net Zero for most of the period only because it has not been scaled by coverage — the trap the exhibit is there to show. The dotted step in 2028 is the illustrative ETS2 extension at the ETS1 price. Right: the allowance price the market would have to reach, month by month, for the effective price to sit on each path — €169 today and €286 by mid-2028 for Net Zero — against the fan's 80 per cent region in money of the day, whose upper edge is €140 at two years.

What NGFS expects is a price on the whole economy, and the step that moves coverage is ETS2. The second trading system is scheduled to become fully operational in 2028 for road transport and buildings — a working assumption of 1,235 Mt of covered emissions (register band 1,150–1,380), all auctioned — which would lift the incidence share from 22 to roughly 63 per cent (band 60–67). As an illustration, the scorecard prices ETS2 at the ETS1 price from January 2028: the probability of being at or above the Below 2 °C path moves from about one per cent to three-quarters at the step, falling back to two-thirds by mid-2028 as the path rises; against Net Zero, from nil to 0.40, then a third. It is an upper bound and a teaching device, not a forecast — ETS2's rules are not settled, its timetable has already been deferred once, and its own reserve releases allowances when its price exceeds €45 in 2020 money, so its price should not be expected to match ETS1's. What the step shows is where the distance to the scenarios actually comes from: coverage first, price second.

Four qualifications are part of the result and travel with any quotation of it. The "at the ETS1 price" condition is strong, and at the legislator's own signalled ceiling the step is smaller and Net Zero falls outside the band. The condition holds at the step and then erodes: the Net Zero path rises by about a third a year in real terms over the fan's two years while the fan's centre is flat, so by 2030 Net Zero would require an allowance near €151 even with ETS2 at full ETS1 price. The NGFS price is not a target but one model's shadow price for a temperature outcome, and other NGFS models put very different prices on the same scenario. And the EU pursues the same targets partly through instruments the model represents as a price, so a low explicit price is not by itself low ambition. The scorecard measures a distance; it does not adjudicate the policy. Every probability is further

conditional on the scenario's non-policy assumptions holding — the EU's carbon cost relative to the rest of the world, the supply-chain build-out rate, and the ETS2 step itself — and a probability quoted without those attached is overstated.

The scorecard scores one dimension of a scenario, and it is not the only one a scenario can fail on.

An NGFS path is a joint statement about prices, technologies, resources and policy, and the carbon price is the dimension this paper can measure against a market. The others can fail on their own, at any carbon price, and the series' work on sustainable aviation fuel shows what that looks like: applied to the same Net Zero 2050 and Below 2 °C paths, Working Papers 03 and 04 found the aggregate SAF mandates physically reachable within the scenarios' feedstock envelopes, but the synthetic-fuel sub-mandate bound by the supply of biogenic CO₂ under Net Zero 2050 through to 2050, hydrogen binding the most abundant advanced pathway first, and the plants actually tracked toward final investment decision an order of magnitude too small for the mandates they serve. None of those constraints is relaxed by a higher allowance price. The reading this paper recommends is therefore joint: the carbon-price distance measured here beside the resource, capacity and policy break points established sector by sector, and a scenario treated as failed when any of them is breached, not only the price.

9. FOR THE READER WITH A BALANCE SHEET

The reader this paper is written for is the finance director or chief financial officer who is responsible for a provision, a capital-allocation decision and a scenario disclosure, and who needs a carbon-cost line that will stand up to an auditor, a board and a supervisor. That is one problem. It has four familiar applications — the investor's risk factor, the lender's covenant test, the insurer's ORSA, the importer's certificate liability — and Working Paper 04's four-lens reading is kept here because each application asks the same distribution a slightly different question. Three properties of the distribution serve all four. Its near-term centre is the market's, so any "view" on the price is a view on a policy surprise or a fuel shock, not on the law, none of whose tested signals improved the forecast. Where the exposure is coal- and gas-fired generation, its effect on emissions is state-dependent through the fuel markets, so the exposure is a function of the gas-coal spread as much as of the allowance; for any other exposure the business supplies its own emissions estimate and applies the price distribution to it. And its width is a deliberately conservative envelope whose historical coverage is reported: a provision set at the upper edge of the central 80 per cent interval would have covered the realised price in every year since 2023, and the interval is wider than those years needed.

9.1 Investors: a risk factor with a measured near-term distribution

Working Paper 01 named carbon a capital risk factor. This paper gives it the first property a risk factor needs: a distribution one can test, at the horizon a portfolio is rebalanced over. The price is modelled as a random walk with fat tails from policy jumps, and over the tested sample nothing forecast it better; a power-generation position's carbon exposure runs through the switching hinge, so the gas-coal spread is part of the position, while an industrial or aviation position's emissions are the holder's own estimate to which the price fan is applied; and the scorecard gives, for any horizon, the probability that Europe's coverage-adjusted carbon price exceeds a given scenario path — the quantity a scenario-weighted valuation needs and usually assumes. For an investment case the marginal basis is the one to read: it is the opportunity cost of a tonne, whether or not the tonne was freely allocated.

9.2 Lenders and project finance: the covenant-test horizon and the fan's

A two-year fan is a trading-horizon object. Loan tenors run far longer, but the debt-service and covenant tests of a project loan are typically assessed on rolling look-back and look-forward windows of one to two years, so the fan spans the horizon over which those tests are usually run. For a carbon-exposed borrower — a generator, steelmaker or airline — the fan gives the distribution of the EUA price over the next eight quarters with a reported historical coverage. That price distribution can be applied to the borrower's independently estimated allowance requirement. The emissions line developed here is narrower: it estimates the response of coal- and gas-fired electricity generation and should not be read as a model of a steelmaker's or airline's abatement. Two uses follow from §7: the fan's width is honest about the tails but wider than the last four years needed, so a lender who used it for a base case would be conservative and should say so; and the fuels — gas, oil and coal together — on their 2022–23 paths from today's level give one named historical stress, run as a path through the generator rather than as a single shock (Figure 21).

9.3 Insurers: what this supports in an ORSA, and what it does not

An ORSA asks two things about climate risk. The first is sensitivity and stress analysis over the business-planning horizon, and the two-year generator answers it directly, complementing the long-term scenario paths rather than replacing them, since the two operate over different decision horizons: a fan with reported historical coverage to 24 months; the power-sector emissions line where the exposure is generation, and the insurer's own emissions estimates elsewhere; named stresses run as paths; and the scorecard's probability by horizon against each NGFS path with the threshold-equivalent price beside it. This is consistent with what EIOPA's 2021 opinion and 2022 application guidance ask of insurers — a materiality assessment and, where material, quantitative scenario analysis proportionate to the exposure; the opinions prescribe neither a method nor a scenario source. The second thing an ORSA asks is the long-term scenario analysis itself, to 2030 and 2050 under at least two warming outcomes. There the model does not replace the NGFS paths; it qualifies them — the coverage correction, the threshold-equivalent price, the break points — and the long-term paths remain given rather than modelled. The planning horizon is usually three to five years and the fan stops at two. It has not been tested beyond twenty-four months; run on to thirty-six, it would carry no test record, and its width would grow as a random walk's — a shape that is honest about what is not known but adds no information (§10).

What the sensitivity-and-stress half looks like in practice is Figure 21, for an illustrative exposure of one million tonnes a year — a net allowance requirement, or a portfolio's look-through — which scales linearly. The deterministic sensitivities an ORSA tabulates (the allowance ± 10 , ± 25 , $+50$ per cent) are each placed on the fan's own probability scale, so the board sees not only what a 25 per cent rise would cost but that the fan gives it roughly one chance in four within a year; two named stresses are run as paths — the allowance's own 2022–23 path against the fan, and the fuels' 2022–23 paths through the generator, which leave the price fan where it is and move the emissions fan; and the two-year cost comes out as a distribution rather than a number.

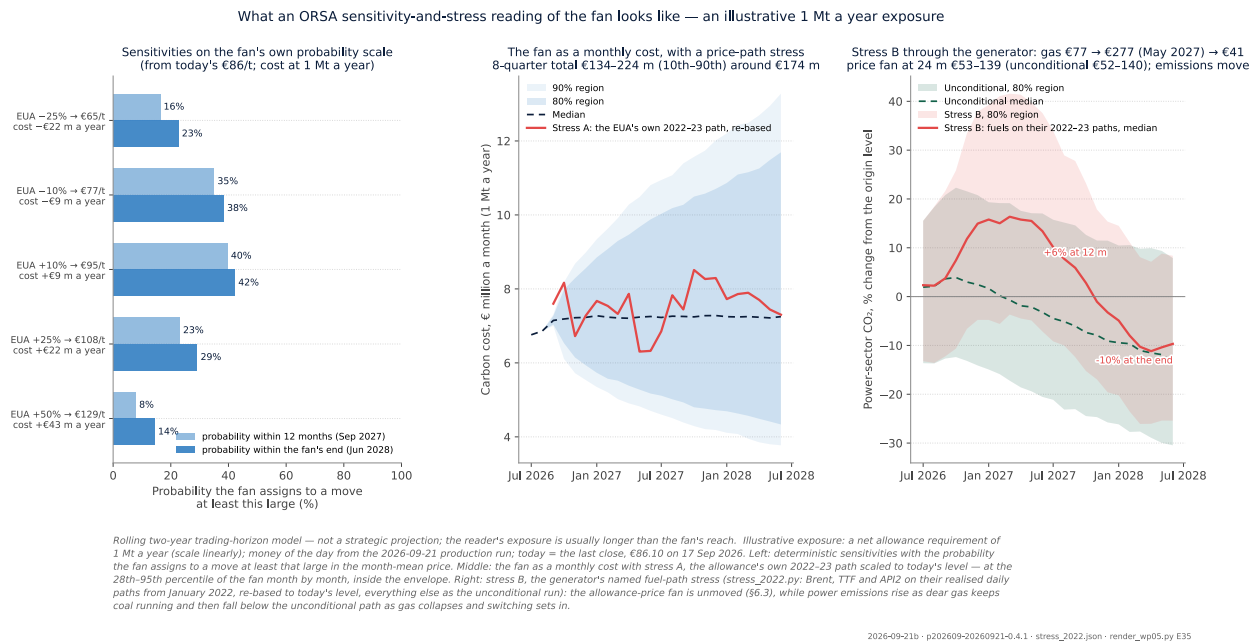


Figure 21 — Left: five deterministic sensitivities on the allowance price from today's €86, each with the cost it implies at one million tonnes a year and the probability the fan assigns to a move at least that large twelve months out and at the fan's end; a 25 per cent rise is more likely than a 25 per cent fall, which is the fan's skew. Middle: the fan as a monthly carbon cost, with stress A — the allowance's own 2022–23 path scaled to today's level, run as a path — at the 28th to 95th percentile of the fan month by month, inside the envelope; the eight-quarter total runs €134–224 million (10th–90th) around €174 million. Right: stress B, the fuel-path stress of \$9.2 run through the generator — Brent, gas and coal on their realised daily paths from January 2022, re-based to today's level, gas going from €77 to €277 by May 2027 and back to €41 by mid-2028. The allowance-price fan does not move (€53–139 at two years against €52–140), as \$6.3 said it would not; power emissions rise 16 per cent above today at the peak of the gas spike, as dear gas keeps coal running, then fall as gas collapses and switching sets in, ending 10 per cent below today against the unconditional path's 13. The stress that matters for the carbon cost is the price path; the stress that matters for emissions is the fuel path.

9.4 Importers: the CBAM certificate liability, priced off the fan

From 2026 an importer of steel, aluminium, cement, fertiliser, hydrogen or electricity carries a liability for the embedded emissions of what it imports, settled in CBAM certificates. That liability is a provision on a balance sheet, and it is a distribution, not a number.

The rule. The number of certificates to surrender for a year is the mass imported times the excess of the good's embedded emissions per tonne over a **free-allocation deduction**, less a reduction in the number of certificates for any carbon price effectively paid in a third country. The deduction is the good's CBAM benchmark times the cross-sectoral correction factor times the **CBAM factor** (Regulation (EU) 2023/956, Articles 22 and 31, implemented by Implementing Regulation (EU) 2025/2620): with actual, verified emissions the benchmark is the process benchmark for the final production step plus its precursors' in the quantities actually consumed (Column A of the Annex); with default values it is the route's default benchmark including precursors (Column B). The CBAM factor (Directive 2003/87/EC, Article 10a(1a)) is 97.5 per cent in 2026, 95 in 2027, 90 in 2028, 77.5 in 2029, 51.5 in 2030, then 39, 26.5, 14 and nil from 2034: it phases out the deduction, so in 2026–27 the importer pays for the excess of its intensity over about 95 per cent of the benchmark, and only by 2034 for the whole embedded emissions. The correction factor is 100 per cent for 2026–2030 (Implementing Decision (EU) 2026/1862). The carbon-price reduction (Article 9, as replaced by Regulation (EU) 2025/2083) applies only where actual emissions are declared, and its conversion —

the price paid divided by the year's average certificate price, times the quantity — is in an implementing act still in draft as this is written. The certificate **price** (Implementing Regulation (EU) 2025/2548) is the volume-weighted average of EU allowance auction clearing prices: for the 2026 obligation the average of the 2026 quarter of importation — €75.36 for the first quarter, €75.28 for the second — and from 2027 the weekly average, applied to certificates bought in that week. Certificates are sold from 1 February 2027; declaration and surrender for a year are due by 30 September of the next; from 2027 the declarant must hold, at each quarter-end, certificates for at least half the year's embedded emissions to date (Regulation (EU) 2025/2083).

The basis. The generator prices the December-benchmark future; the auction clears below it. On 1,239 auction days since 2021 the volume-weighted clearing price sat 1.4 per cent (€1.07) below the same-day close, and the difference is a carry to December: about -2.9 per cent in the first quarter of a year, -2.3 in the second, -1.3 in the third, -0.7 in the fourth. Applied to the platform's auction table, the 2025/2548 rule reproduces the Commission's two published 2026 prices to the cent, so the translation from the fan to a certificate price is a measured, seasonal discount, an order of magnitude smaller than the fan's width.

The case. An importer brings in 10,000 tonnes a quarter of continuously cast slab of non-alloy steel (CN 7207 12 10) made by the blast-furnace-basic-oxygen route in a third country where no carbon price is effectively paid, and declares actual, verified embedded emissions of 2.0 tCO₂ per tonne. With the installation's hot-metal input at 1.08 tonnes per tonne of slab, the deduction's benchmark is the slab process benchmark of 0.188 plus 1.08 times the hot-metal benchmark of 1.089, which is 1.364 tCO₂ per tonne — the same figure as the route's default benchmark, so an installation charging more scrap, and therefore consuming less hot metal per tonne, would have a smaller deduction; whether it owed more or less would depend on how far its own actual embedded emissions fell with the scrap charge, which the actual-data declaration also captures. The certificates due are 6,701 a quarter in 2026 (34 per cent of the 20,000 tonnes embedded), 7,042 in 2027, 7,724 in 2028, 12,975 in 2030 and 20,000 from 2034. Two consequences follow from the price rule. On the day this is written the first two quarters of the 2026 liability are fixed at the published prices (€505 k and €504 k), the third is near-final at the auction average to 25 September (€82.09, €550 k) and will settle within days, and only the fourth is priced from the fan — a certificate price of €71–100 (5th–95th) around €86 after the fourth-quarter basis — so the year's total runs **€2.03–2.23 million around €2.13 million**, an uncertainty of about ±5 per cent. And from 2027 the price is the weekly average in the week of purchase, so the liability is a purchase-timing problem and the right object is the path-wise mean of the fan over the purchase window, not a quarterly quantile. The distribution below assumes purchases spread evenly through the year; the quarter-end holding rule requires only that certificates for at least half the year-to-date embedded emissions be held, so it bounds how late purchases can be left without requiring them to be even, and front- or back-loading within that bound is a timing exposure the fan can price but this figure does not. On that assumption, 28,168 certificates at a mean price of €58–122 (5th–95th) around €86 give **€1.6–3.4 million around €2.4 million**, and — for the exposure the importer will carry once the deduction has gone — 80,000 certificates give €4.6–9.8 million around €6.9 million. Here the uncertainty is of the order of the expected value within the year, for 2027 rather than 2026 and for the certificates actually due rather than the embedded emissions. The point for an importer — and for the CBAM service providers now building around them — is that the provision is dominated by the allowance price once the count is known, and that the fan, rerun every morning and translated through the measured basis, is the instrument that keeps it current. Figure 22 draws the case.

The CBAM certificate liability of a worked importer, priced off the fan: 10,000 t of BF-BOF slab a quarter at 2.0 tCO₂/t

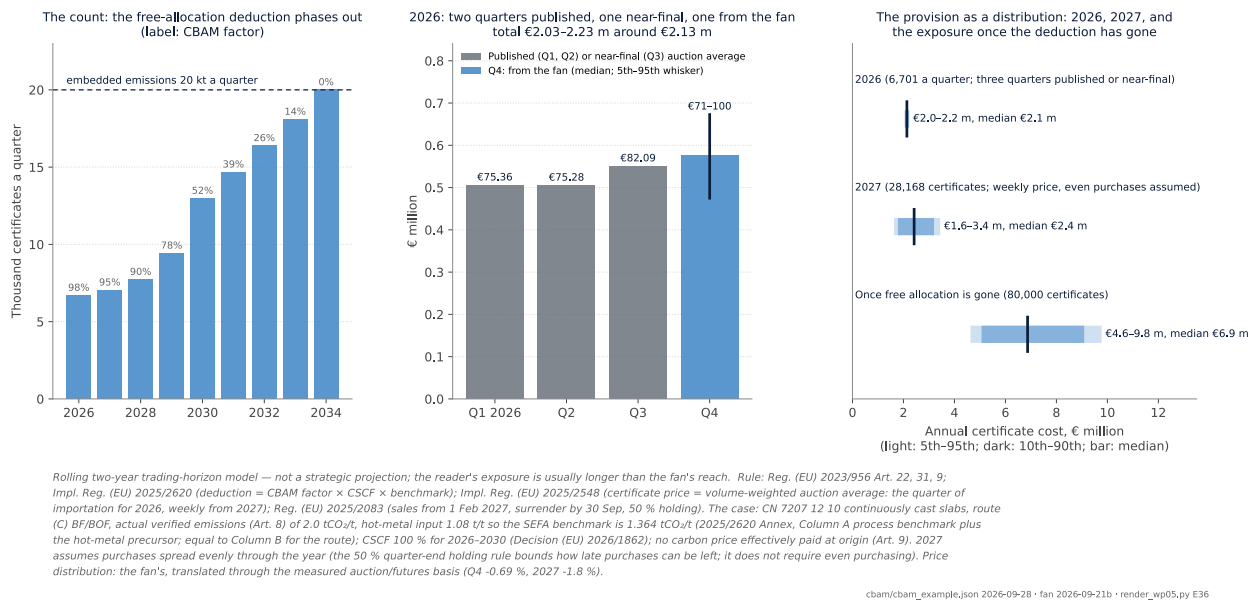


Figure 22 — Left: the certificates the importer owes per quarter, 2026 to 2034, against its 20,000 tonnes of embedded emissions; the label on each bar is the CBAM factor, which scales the free-allocation deduction, so the count rises from 34 per cent of embedded emissions to all of it as the deduction phases out. Centre: the 2026 provision quarter by quarter — two quarters at the published prices, the third at the near-final auction average, the fourth from the fan with its 5th–95th whisker — totalling €2.03–2.23 million around €2.13 million. Right: the annual liability as a distribution for 2026, for 2027 at the weekly price in the week of purchase assuming purchases spread evenly through the year, and for the exposure once the deduction has gone. The case is a BF-BOF slab declared on actual data with the benchmark and correction factor read from the acts; the price distributions are the fan's.

9.5 The shared exposure

As in Working Paper 04, the four readings are faces of one thing: the near-term carbon cost behaves, over the tested sample, like a random walk whose tails are policy jumps and market shocks, whose effect on emissions is set by the gas market, and whose distance from the scenarios supervisors use is a matter of coverage before it is a matter of price. That is worth naming once rather than four times.

10. LIMITATIONS, AND WHAT COMES NEXT

What the model does not do. It does not know where the price level should be beyond the market's own read; its centre is flat by construction and its width is history's. It does not model the long horizon: the structural layer sketched in Working Paper 01 — cap, bank, reserve, ETS2 as priced blocks — is specified alongside this work and parked until the bank ledger the platform is building closes the identity of §3; the box below says how the two fit together, and the next paper carries the implementation. Weather enters as climatology rather than as anomaly draws, and pending legislative dossiers are not yet drawn as probabilities of enactment; both are specified and neither is admitted. ETS2 is an illustration, not a priced block. Aviation and maritime are in the cap and not in the emissions-response model; no aviation emissions response is inferred in this paper (§5). The emissions equation's trend is estimated from 2019 and fixed; its under-coverage on the 2024–25 origins says a time-varying trend is the next thing to test.

Beyond twenty-four months: a different regime, not a longer forecast

A reader with a three- or five-year horizon will ask why the fan stops at two years and whether it can simply be run on. It cannot, and the reason is the paper's own finding. Over one to twenty-four months the price behaved, in the tested sample, indistinguishably from a martingale plus surprises: the market owns the level, history calibrates the width, and no tested model forecast the price better than a random walk with tails drawn from the record. Run that model to 2030 and its centre stays flat while its width grows with the square root of time — a shape that says nothing about the one thing that decides the long-run level, which is that the cap reaches zero around 2039, the bank of allowances held by participants is finite, and the price must eventually be whatever makes cumulative emissions fit cumulative supply. The long horizon is therefore a second regime of the same framework, answered by a different model (Figure 23).

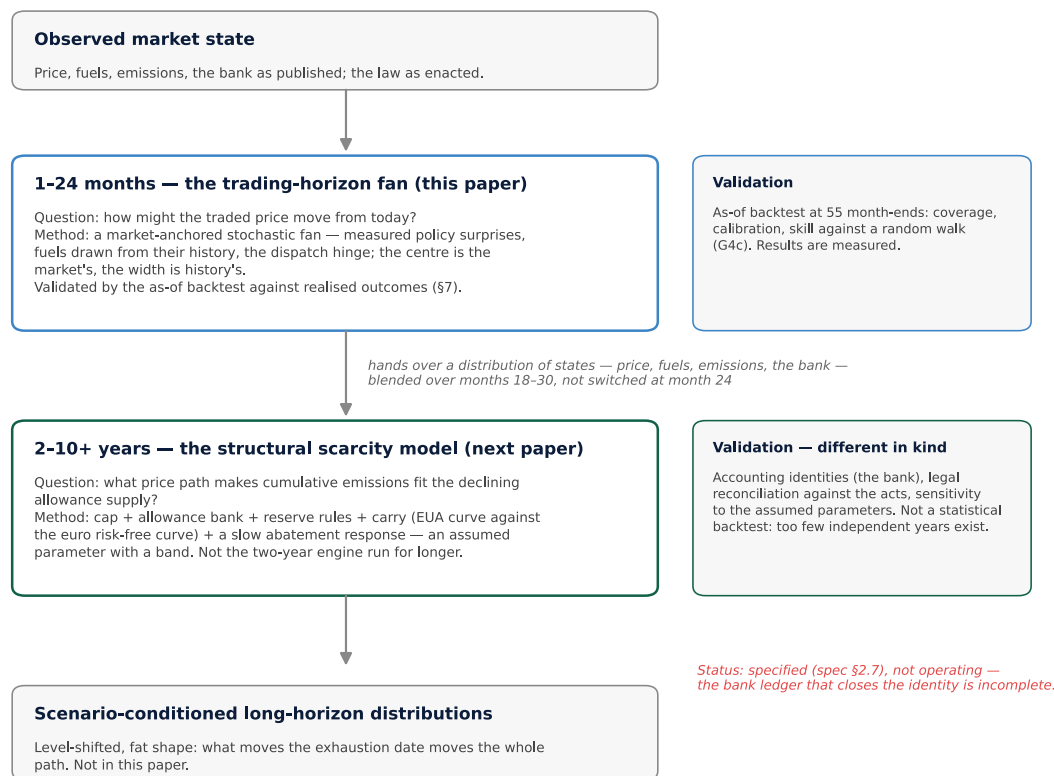
Horizon	Central question	Method	Validation
1–24 months	How might the traded price move from today?	Market-anchored stochastic fan: measured policy surprises, fuels and dispatch	As-of backtest against realised outcomes (§7)
2–10+ years	What price path makes cumulative emissions fit the declining allowance supply?	Structural cap–bank–reserve–abatement model	Accounting identities, legal reconciliation, parameter sensitivities — not a statistical backtest

The handover. The second model is not the first run for longer. At the join, the two-year simulation supplies a distribution of starting states — price, fuels, emissions and, once the ledger closes, the allowance bank — and the structural model evolves those states under the cap, the reserve's rules and the abatement assumptions. The join is more likely a blending period over months eighteen to thirty than a switch at exactly month twenty-four, since neither regime is sharply right at the boundary; that design is open.

What determines the longer path. Four things, two known and two assumed. The cap and the reserve's rules are written in law and reconciled in Working Paper 02. The allowance bank, once the platform's ledger closes the identity of §3, determines when cumulative supply binds. The carry — the rate at which the expected price grows while the bank is positive — is measured: the EUA December futures curve gives the market's total carry, and the AAA euro-area spot curve with the €STR series separates its risk-free component from the premium for holding carbon. The slow abatement response — emissions per euro of price over years — is not identified on the data available and enters as an assumed parameter with a band. The bank ledger and the abatement response matter far more to the long-run level than any of the econometrics in this paper.

Why it is not yet operating, and how it will be validated. The bank identity does not close from the components on hand (§3); until the ledger closes it, the model has no starting bank and is parked. When it runs it will not be validated the way the fan was — seven annual observations allow no credible statistical backtest of a ten-year path — but by construction: accounting identities, reconciliation of every legal parameter against the acts, and sensitivity to each assumed parameter, with the assumptions in view. No provisional five- or ten-year fan is given here, because the paper's credibility rests on the line between measured results and assumptions. Until the structural layer is complete the NGFS paths remain the external long-term scenarios, read with §8's qualifications.

Two regimes, one framework: the tested fan is the near end of a structural model, not its whole



Architecture only — no long-horizon result is claimed in Working Paper 05.

spec §2.7 (ruled 2026-09-19) · review 2026-09-28 · render_wp05.py E34

Figure 23 — The architecture, without results. The tested two-year fan takes the observed market state and hands a distribution of states to the structural scarcity model, which evolves them under the cap, the bank, the reserve's rules, the measured carry and the assumed abatement response; the two are validated differently, and the second is specified but not operating.

What the numbers are conditional on. Every backtest number rests on 56 overlapping month-ends and, at long leads, a handful of independent windows; the by-year cuts show which years earned which result. Every scorecard probability rests on the coverage shares of 2024 carried forward, on an EU-15 scenario path read against EU27 shares, and on the scenario's non-policy assumptions holding. Every result carries the data version stamped on the rows it came from, and the publication manifest keeps three kinds of result apart. The generator's results — the fan, the backtests, the stresses — are generated from a frozen data vintage and regenerate from it; the research analyses of §3–§5 and the auction basis of §9.4 are dated extracts of the platform's database, reproducible from the extract as committed but not from the vintage; and the legal inputs of §9.4 — benchmarks, factors, published certificate prices — are typed from the acts and verified against them, not generated at all.

What comes next. The forward record of §7.2, which accrues from the platform's daily runs and will be scored once the scoring job runs. The CBAM carbon-price reduction re-read once the Commission's implementing act is adopted. Working Paper 06 reads the SAF mandates of Working Paper 04 against the compliance data now accumulating; the long-horizon structural model of the box above, implemented and reconciled, is a paper of its own. And, from this paper's readers, corrections: the event register's recall gap, the dispatch constants for a generating stock that is changing.

FURTHER MATERIAL

Everything behind this paper is kept and is available on request. The methods note gives every formula used, under the numbering the project keeps: the event register and the surprise construction with its two denominators; the long-sample proxy VAR and the local projections; the switching hinge and its bootstrap; the price-to-emissions composition; the monthly engine, its prior and its guard; the daily layer and the bridge; the fuel processes; the emissions equation and its calibration factor; the effective carbon price and the scorecard; the consolidated backtest with its tests explained and the complete results table; and the price calibration factor with its figure. The claim register lists every number this paper asserts with the result file and expression that yields it, the value that file holds at the publication commit and its caveat, and is rebuilt on every build; a companion register lists the paper's legal, perimeter and scope statements with the source each was checked against. The exhibit register maps each figure to the script that draws it and the data version it was drawn from; every figure carries that stamp. The data versions, parameter sets and code versions of every result are recorded in a publication manifest that is verified at the publishing commit and states, for each result file, whether it regenerates from the frozen vintage, only from a dated research extract, or not at all. Write to the author for any of these.

SOURCES

Working papers in this series (peterkeutgens.com/writing). Keutgens, P. (2026a). *Carbon as a Capital Risk Factor*. Working Paper 01, February 2026. — Keutgens, P. (2026b). *The EU ETS in Numbers*. Working Paper 02, June 2026; v0.3, July 2026. — Keutgens, P. (2026c). *Sustainable Aviation Fuel — From Feedstock to Fuel*. Working Paper 03, June 2026. — Keutgens, P. (2026d). *Sustainable Aviation Fuel — From Network to Scenario*. Working Paper 04, July 2026.

Econometrics and forecasting. Känzig, D. R. (2023). "The unequal economic consequences of carbon pricing." *American Economic Review* 113(6), 1409–1455. — Bauer, M. D., Känzig, D. R. and Rudebusch, G. D. (2026). "Carbon pricing and inflation expectations." CEPR Discussion Paper DP21355, April 2026 (also Hutchins Center Working Paper 106, Brookings Institution). The extended carbon-policy surprise series to December 2024 is released by the authors under CC BY 4.0 (github.com/dkaenzig/carbon-policy-surprises-update) and held in the platform's document store as release `bkr-2026`. — Working, H. (1960). "Note on the correlation of first differences of averages in a random chain." *Econometrica* 28(4), 916–918. — Diebold, F. X. and Mariano, R. S. (1995). "Comparing predictive accuracy." *Journal of Business & Economic Statistics* 13(3), 253–263. — Gneiting, T. and Raftery, A. E. (2007). "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102(477), 359–378. — Jordà, Ò. (2005). "Estimation and inference of impulse responses by local projections." *American Economic Review* 95(1), 161–182. — Stock, J. H. and Watson, M. W. (2018). "Identification and estimation of dynamic causal effects in macroeconomics using external instruments." *Economic Journal* 128(610), 917–948. — Anderson, T. W. and Rubin, H. (1949). "Estimation of the parameters of a single equation in a complete system of stochastic equations." *Annals of Mathematical Statistics* 20(1), 46–63. — Kilian, L. (1998). "Small-sample confidence intervals for impulse response functions." *Review of Economics and Statistics* 80(2), 218–230. — Litterman, R. B. (1986). "Forecasting with Bayesian vector autoregressions — five years of experience." *Journal of Business & Economic Statistics* 4(1), 25–38. — Newey, W. K. and West, K. D. (1987). "A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica* 55(3), 703–708.

Scenarios and supervision. Network for Greening the Financial System (2024/25). *NGFS Climate Scenarios*, Phase 5 (v5.1): scenario data and technical documentation; GCAM 6.0 NGFS. — EIOPA (2021). *Opinion on the supervision of the use of climate change risk scenarios in ORSA*. — EIOPA (2022). *Application guidance on climate change materiality assessments and climate change scenarios in ORSA*.

Legal acts (EUR-Lex, cited by article). Directive 2003/87/EC establishing a system for greenhouse gas emission allowance trading, as amended, in particular Article 10a(1a) and (5) as amended by Directive (EU) 2023/959. —

Decision (EU) 2015/1814 concerning the establishment and operation of a market stability reserve, as amended. — Regulation (EU) 2023/956 establishing a carbon border adjustment mechanism, as amended by Regulation (EU) 2025/2083 (simplification; Articles 6, 9, 20 and 22 as replaced). — Commission Implementing Regulation (EU) 2025/2548 (the price of CBAM certificates). — Commission Implementing Regulation (EU) 2025/2620 of 16 December 2025 (adjustment for free allocation; Annex point 5, CBAM benchmarks). — Commission Implementing Regulation (EU) 2025/2621 of 16 December 2025 (default values for embedded emissions). — Commission Implementing Decision (EU) 2026/1862 of 23 July 2026 determining the uniform cross-sectoral correction factor for 2026 to 2030. — European Commission, draft implementing regulation on the carbon price paid in a third country (Ares(2026)4841230, 13 May 2026; not adopted at the date of writing). — European Commission, "Price of CBAM certificates", publications of 7 April and 6 July 2026.

Data. The platform's document store and database: EEX auction results; ICE EUA December futures, TTF, Brent and API2 settlements (deduplicated through one shared view); ENTSO-E generation and the platform's power-sector emissions series; Eurostat industrial production; the Union Registry and the Commission's TNAC publications; the regulatory event register described in §4; ERA5 2 m temperature, from which the platform's heating- and cooling-degree-day anomalies are built — Copernicus Climate Change Service (C3S), Climate Data Store: *ERA5 post-processed daily statistics on single levels from 1940 to present*, DOI 10.24381/cds.4991cf48. Contains modified Copernicus Climate Change Service information 2026; neither the European Commission nor ECMWF is responsible for any use that may be made of it. Every figure carries the data version it was drawn from.

The views expressed are solely those of the author and do not represent any affiliated institution. This paper is part of ongoing independent research and should not be interpreted as investment, actuarial, legal or regulatory advice.