

Policy-State Gated Regime Dynamics in European Emission Allowance Futures

wanyok Attisattapong

Thammasat University

Pasin Marupanthorn

quantfilab@gmail.com

Maejo University

Research Article

Keywords: carbon allowances, EU ETS, regime switching, policy design, energy-market stress, market activity

Posted Date: July 1st, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9966945/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Discover Sustainability on September 13th, 2026. See the published version at <https://doi.org/10.1007/s43621-026-04695-4>.

Policy-State Gated Regime Dynamics in European Emission Allowance Futures

Wanyok Attisattapong¹ and Pasin Marupanthorn^{2*}

¹Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University, Pathum Thani, 12120, Thailand.

^{2*}Department of Mathematics, Faculty of Science, Maejo University, Chiang Mai, 50290, Thailand.

*Corresponding author(s). E-mail(s): quantfilab@gmail.com;
Contributing authors: wanyok@mathstat.sci.tu.ac.th;

Abstract

Emission allowances are policy-created financial assets whose return dynamics depend on regulatory scarcity, market surplus, energy-market stress, and trading activity. This paper estimates a two-state time-varying transition probability Markov model for European Union Allowance futures returns using Refinitiv Workspace data. The transition law depends on four observable variables: a market-implied EUA price-state proxy, official TNAC, Dutch TTF gas volatility, and an EUA futures activity proxy. The gate is aligned to the next trading day's return, and TNAC values enter only from their public release dates onward. On the common estimation sample from May 15, 2019 to May 28, 2026, the full gate improves fit within the Gaussian Markov family by log likelihood and AIC, but not by BIC, and heavy-tailed GARCH-family benchmarks remain stronger by both AIC and BIC. Common-sample one-step-ahead ranking results are modestly favorable to the Markov score, but fixed train-validation-test splits with tail thresholds fixed from the training sample do not establish real-time forecasting dominance over GARCH, and post-crisis VaR calibration remains weak. The contribution is therefore interpretive and associative: observable policy surplus, energy volatility, and activity variables help explain variation in EUA regime-transition probabilities, while heavy-tailed volatility models remain important density-forecasting benchmarks.

Keywords: carbon allowances, EU ETS, regime switching, policy design, energy-market stress, market activity

1 Introduction and literature review

Direct carbon pricing is now a material macro-financial institution, not merely an environmental-policy instrument. The World Bank Group reports that direct carbon pricing covered nearly 30 percent of global greenhouse-gas emissions across 87 implemented policies and mobilized more than USD 107 billion of public revenue in 2025 (World Bank Group 2026). Within that architecture, the European Union Emissions Trading System (EU ETS) remains the most important laboratory for stochastic-process research because it combines a liquid exchange-traded allowance market with a legally engineered supply schedule, intertemporal banking, auctioning, compliance demand, and repeated policy intervention. Its allowance price is therefore neither a standard commodity price nor a purely financial asset price. It is the market value of a policy-created compliance right whose scarcity is jointly determined by the cap, banking incentives, fuel-switching margins, regulatory credibility, and the expectations of both compliance and financial traders.

The institutional environment has also changed materially in the most recent phase of the EU ETS. The 2023 reform under Directive (EU) 2023/959 strengthened the cap trajectory, revised free-allocation and Market Stability Reserve (MSR) rules, expanded the system to maritime transport, and created a separate ETS2 for buildings and road transport (European Parliament and Council of the European Union 2023; European Commission 2026a). These reforms matter for return modeling because they make the allowance market more forward looking: as the cap tightens, market participants price not only current emissions demand but also the credibility and timing of future scarcity. Pahle et al. (2025) describe this setting as the emerging “endgame” of the EU ETS, in which the path toward climate neutrality and a near-zero long-run cap may generate new frictions in allowance banking, liquidity, expectations, and price formation. Consequently, the empirical object in a carbon-market stochastic process is not simply volatility persistence; it is the law of motion of returns under a changing market design.

The Market Stability Reserve is central to this problem. It adjusts auction supply in response to the total number of allowances in circulation (TNAC), and the European Commission publishes the TNAC annually as the key surplus indicator used for MSR operation (European Commission 2025, 2026b). The MSR was designed to absorb historical surplus and increase resilience to demand shocks, but the literature shows that it can stabilize or destabilize the market depending on banking behavior, expectations, and the interaction between quantity rules and price formation (Perino et al. 2022; Borghesi et al. 2023). This is the first reason why a regime model for EUA returns should not be purely latent. If the policy rule changes future allowance supply in response to an observable surplus measure, then the transition between low- and high-stress return states should plausibly depend on observable policy-state variables.

A second reason is that EUA returns are tightly connected to short-run energy-market economics. The allowance price is influenced by the relative cost of coal- and gas-fired generation because fuel switching changes emissions demand and marginal abatement costs. Bai and Okullo (2023) show that the EU ETS allowance price responds systematically to gas and coal prices and that allowance-cost pass-through in major European power markets is economically substantial. This matters for the

present paper because the 2021–2022 European energy crisis made gas-market volatility an explicit stress channel for carbon prices. A model that ignores gas-market turbulence can still fit volatility *ex post*, but it cannot explain whether stress-state transitions are being triggered by energy-market pressure that is observable before the next return is realized.

A third reason is the increasing financialization of carbon allowances. EUA futures are used by compliance firms, utilities, commodity traders, asset managers, and risk managers. [Demiralay et al. \(2022\)](#) treat carbon futures as an emerging asset class and show that they can provide hedging and diversification value, although the hedging relationship is time varying and deteriorated during the COVID-19 episode. More recently, [Wei et al. \(2025\)](#) examine the financialization of the EU ETS and show that external financial and commodity-market factors affect the system differently across quantiles. This literature motivates the inclusion of market-activity variables such as EUA futures volume. Trading activity is not only a liquidity control; it can proxy the intensity with which information, hedging demand, and speculative pressure are incorporated into the allowance market. For that reason, volume may influence regime persistence even when it is not the primary source of a price shock.

The econometric literature on carbon prices has long documented that allowance returns are non-Gaussian, heteroskedastic, and regime dependent. [Seifert et al. \(2008\)](#) show that carbon prices behave differently from standard commodities. [Benz and Trueck \(2009\)](#) and [Daskalakis et al. \(2009\)](#) document non-normality, jumps, and derivative-pricing implications. [Gronwald and Ketterer \(2012\)](#) link large price moves to jump-like dynamics and news, while [Li et al. \(2016\)](#) demonstrate the relevance of regime-switching jump diffusion for allowance derivatives. This stream establishes why a single-regime Gaussian diffusion is an inadequate benchmark, but it also creates a higher burden for new work. Since heavy tails, jumps, and volatility clustering are already known facts, the novelty of a new carbon-market process cannot come from rediscovering them.

The methodological distinction is therefore important. Markov-switching models provide a natural way to represent unobserved shifts in return distributions ([Hamilton 1989](#)). Time-varying transition probability models extend this idea by allowing transition probabilities to depend on observable state variables ([Filardo 1994](#)). In the EU ETS setting, this distinction is economically meaningful. A latent Markov model can label high- and low-volatility states after observing returns, but it does not tell the reader whether those states are connected to policy surplus, energy-market stress, or trading activity. A policy-state gated Markov process instead asks whether these observables alter the probability of remaining in or moving into the high-stress state. The transition law, rather than only the conditional variance, becomes the object of economic interpretation.

Recent work reinforces the importance of interpretable state variables. [Fuchs et al. \(2024\)](#) construct a market-based measure of carbon price uncertainty from EUA options and show that uncertainty is high, persistent, and affected by major climate-policy events. [Cepni et al. \(2025\)](#) further show that the persistence of carbon-price uncertainty varies over time and is linked to carbon-policy uncertainty. [Eksi et al. \(2025\)](#) estimate a partial-information regime-switching model for EU ETS prices

through October 2024, confirming that richer state inference is useful in allowance markets. These modern contributions strengthen the case for regime-based modeling and highlight an unresolved issue: the transition specification itself should be explicitly connected to the policy and market variables that define the EU ETS.

This paper addresses that gap by estimating a two-regime time-varying transition probability model for EUA futures returns. The transition law depends on four observable state variables motivated by EU ETS economics and market functioning: a market-implied EUA price-state proxy, official TNAC, Dutch TTF gas volatility, and an EUA futures activity proxy. The contribution is interpretive rather than causal. First, the paper places policy-related and market-activity variables directly in the transition equation instead of treating them only as *ex post* narratives for latent states. Second, it estimates the model on a Refinitiv EUA futures panel rather than on public ETF proxies. Third, it evaluates the specification against latent-regime, nested-gate, and stronger volatility benchmarks, including Student- t GARCH, GJR-GARCH, and EGARCH, while separating common-sample regime interpretation from held-out validation. Fourth, it translates coefficient estimates into scenario-based changes in transition probabilities and stress-state occupancy, with accompanying uncertainty intervals.

This framing deliberately avoids an exaggerated claim. The purpose of the model is not to impose a causal interpretation on TNAC, gas volatility, or trading activity, nor to claim automatic dominance over every volatility benchmark. It is to test whether these observable variables are associated with economically interpretable variation in the timing and persistence of return regimes. If they are, then the model contributes a policy-state interpretation that a reduced-form volatility recursion cannot provide.

The remainder of the manuscript proceeds as follows. Section 2 states the economic channels and hypotheses. Section 3 develops the stochastic process. Section 4 describes estimation and forecast evaluation. Section 5 presents the data construction and empirical sample. Section 6 reports the empirical results. Section 7 concludes.

2 Economic channels and hypotheses

Emission allowances are regulatory compliance rights, not ordinary commodities. Their price is formed by the interaction of a legally fixed supply path, compliance demand from covered installations, intertemporal banking, fuel-switching incentives, and the trading behavior of financial intermediaries. This institutional structure makes the conditional distribution of EUA returns inherently state dependent. When the market expects future scarcity, a small change in policy credibility, energy costs, or hedging demand can move the allowance price sharply. When the market is perceived as well supplied, the same shock may be absorbed by banked allowances and produce a weaker return response. The relevant economic object is therefore not only the level of volatility, but the probability that the market moves into, remains in, or exits a stress state.

The policy channel is centered on the cap, auction supply, allowance banking, and the Market Stability Reserve (MSR). The total number of allowances in circulation (TNAC) is the official surplus measure used to guide MSR operation, while auction

volumes and cancellation rules affect the expected quantity of allowances available to the market (European Commission 2025, 2026b; Perino et al. 2022; Borghesi et al. 2023). These rules create a feedback channel from observed surplus to expected future supply. A high TNAC may signal weak current scarcity, but it may also activate expectations of future supply withdrawal through the MSR. Conversely, a low TNAC may imply immediate scarcity but less future MSR intake. Hence the expected sign of the TNAC effect on return regimes is not imposed mechanically; it is an empirical question about how the market prices surplus, banking value, and future policy adjustment.

Energy-market conditions provide the second channel. EUA demand is linked to the marginal emissions intensity of power generation because gas, coal, and power prices affect fuel switching and short-run abatement incentives (Bai and Okullo 2023). In normal periods, changes in fuel prices may be reflected gradually through expected compliance demand. During energy stress, however, gas-price volatility can transmit rapidly to EUA returns because it changes both the expected use of fossil generation and the uncertainty surrounding marginal abatement costs. Dutch TTF gas volatility is therefore interpreted as an observable energy-stress variable rather than as a generic commodity control.

Market activity provides the third channel. EUA futures volume reflects the intensity of compliance hedging, speculative trading, portfolio rebalancing, and information incorporation in a financialized carbon market (Demiralay et al. 2022; Wei et al. 2025). High volume need not imply a one-day shock; it can instead indicate that many participants are updating positions under common policy or energy information. For this reason, trading activity may affect regime persistence by making the high-stress state more self-sustaining, even after the initial shock has passed.

Let B_t denote the policy-surplus state, represented empirically by TNAC and related surplus information; let A_t denote auction-supply and MSR-related supply conditions; let E_t denote energy-market conditions; let D_t denote policy-event information; and let M_t denote market-activity and macro-financial controls. A compact state vector is

$$Y_t = (B_t, A_t, E_t, D_t, M_t)'. \quad (1)$$

In the empirical specification, this broad vector is represented by four observable transition-gate variables: a rolling EUA price-state proxy, official TNAC, 21-day annualized Dutch TTF gas volatility, and EUA futures log volume. The price-state proxy summarizes the relative position of the allowance price, TNAC captures the official policy-surplus state, gas volatility captures energy-market stress, and log volume captures the intensity of trading and information processing.

The empirical channel summary is as follows. A low-stress regime is expected when surplus conditions are stable, energy volatility is moderate, and trading activity is consistent with ordinary hedging and liquidity provision. A high-stress regime is expected when policy-surplus information, energy-market turbulence, or trading activity changes the market’s assessment of future scarcity. The proposed model does not require these channels to move returns in the same direction; it requires only that they alter the transition probabilities between return states.

Four hypotheses organize the empirical design.

H1: Regime dependence. EUA returns are sufficiently heavy-tailed and state dependent that a single-regime Gaussian return model is an inadequate benchmark.

H2: Policy-state association. Official policy-state information, especially TNAC and the market-implied EUA price-state proxy, is associated with changes in the probabilities of entering and exiting the high-stress regime.

H3: Energy and activity channels. TTF gas volatility and the EUA activity proxy add incremental transition information beyond the market-implied price-state proxy and TNAC alone.

H4: Diagnostic value. A gated Markov process improves the interpretation and classification of EUA return regimes relative to a purely latent Markov model and simpler nested gates, even if heavy-tailed volatility models remain competitive for reduced-form density fit.

These hypotheses deliberately separate density fitting from economic state identification. The purpose of the proposed process is not to replace every volatility benchmark, but to test whether observable policy, energy, and market-activity variables help explain the law of motion of EUA return regimes.

3 Policy-regime stochastic process

Let S_t denote the EUA futures price, $X_t = \log S_t$, and let $r_{t+1} = 100(X_{t+1} - X_t)$ be the daily log return measured in percent. The economic channels in Section 2 suggest a regime-dependent return process rather than a single invariant diffusion. Because the contribution of the paper is empirical and discrete-time, the estimated specification begins directly with a state-dependent daily observation equation rather than with a continuous-time jump-diffusion representation.

Following the logic of Markov-switching and time-varying transition probability models (Hamilton 1989; Filardo 1994), the baseline daily model is a two-state Markov normal process with state-dependent conditional moments,

$$r_{t+1} \mid (Z_{t+1} = j, \mathcal{F}_t) \sim N(\mu_j, \sigma_j^2), \quad j \in \{0, 1\}, \quad (2)$$

where

$$\mathcal{F}_t = \sigma(r_s, G_s : s \leq t)$$

is the information set observed by the end of trading day t . In the empirical application, daily gate variables are treated as end-of-day observables, so the forecast target is always the next close-to-close return r_{t+1} rather than the same-day return r_t . The states are labelled by volatility, with $\sigma_0 < \sigma_1$; state 0 is therefore interpreted as the low-stress regime and state 1 as the high-stress regime. This labelling restriction is important because the latent-state likelihood is otherwise invariant to a relabelling of the two regimes.

Let the active gate be

$$G_t = (\tilde{s}_t, \tilde{\tau}_t, \tilde{g}_t, \tilde{v}_t)', \quad (3)$$

where $\tilde{s}_t$ is the standardized market-implied price-state proxy, $\tilde{\tau}_t$ is standardized TNAC, $\tilde{g}_t$ is standardized Dutch TTF gas volatility, and $\tilde{v}_t$ is the standardized EUA

activity proxy. The transition law is

$$p_{i0,t} \equiv \Pr(Z_{t+1} = 0 \mid Z_t = i, G_t) = \Lambda\left(a_{i0} + b_{i0}^{(s)} \tilde{s}_t + b_{i0}^{(\tau)} \tilde{\tau}_t + b_{i0}^{(g)} \tilde{g}_t + b_{i0}^{(v)} \tilde{v}_t\right), \quad i \in \{0, 1\}, \quad (4)$$

where $\Lambda(x) = (1 + e^{-x})^{-1}$. Since there are two states, $p_{i1,t} = 1 - p_{i0,t}$ and the time- t transition matrix is

$$P_t(G_t) = \begin{pmatrix} p_{00,t} & 1 - p_{00,t} \\ p_{10,t} & 1 - p_{10,t} \end{pmatrix}. \quad (5)$$

The coefficients have direct economic meanings because all gate variables are standardized. In the first row, a negative coefficient lowers low-state persistence $p_{00,t}$ and raises the probability of entering the high-stress state. In the second row, a negative coefficient lowers the exit probability $p_{10,t}$ and therefore increases high-state persistence. This is the key difference from a purely latent Markov model: the transition probabilities are not fixed constants inferred only from returns, but functions of observable policy surplus, energy stress, and trading activity.

The filtering recursion gives the likelihood and the state probabilities used in the empirical section. If $\pi_{i,t} = \Pr(Z_t = i \mid \mathcal{F}_t)$ denotes the filtered probability, the one-step predicted probability for state j is

$$\xi_{j,t+1|t} = \Pr(Z_{t+1} = j \mid \mathcal{F}_t) = \sum_{i=0}^1 \pi_{i,t} p_{ij,t}. \quad (6)$$

After observing r_{t+1} , Bayes' rule updates the filtered probability as

$$\pi_{j,t+1} = \frac{\xi_{j,t+1|t} \phi(r_{t+1}; \mu_j, \sigma_j^2)}{\sum_{\ell=0}^1 \xi_{\ell,t+1|t} \phi(r_{t+1}; \mu_\ell, \sigma_\ell^2)}, \quad (7)$$

where $\phi(\cdot; \mu, \sigma^2)$ is the normal density. Equations (6)–(7) show how covariate-gated transition probabilities and realized returns jointly determine the high-stress probability plotted and evaluated later.

The one-day predictive density is the normal mixture

$$p(r_{t+1} \mid \mathcal{F}_t) = \sum_{j=0}^1 \xi_{j,t+1|t} \phi(r_{t+1}; \mu_j, \sigma_j^2). \quad (8)$$

Thus the predictive mean and variance are

$$m_{t+1} = \sum_{j=0}^1 \xi_{j,t+1|t} \mu_j, \quad v_{t+1} = \sum_{j=0}^1 \xi_{j,t+1|t} (\sigma_j^2 + \mu_j^2) - m_{t+1}^2. \quad (9)$$

For a left-tail probability level α , the predictive quantile $q_{\alpha,t+1}$ solves

$$\sum_{j=0}^1 \xi_{j,t+1|t} \Phi\left(\frac{q_{\alpha,t+1} - \mu_j}{\sigma_j}\right) = \alpha, \quad (10)$$

where $\Phi(\cdot)$ is the standard normal distribution function. The corresponding expected shortfall of returns is

$$ES_{\alpha,t+1} = \frac{1}{\alpha} \sum_{j=0}^1 \xi_{j,t+1|t} [\mu_j \Phi(z_{j,t+1}) - \sigma_j \phi(z_{j,t+1})], \quad z_{j,t+1} = \frac{q_{\alpha,t+1} - \mu_j}{\sigma_j}. \quad (11)$$

These formulas explain why the proposed model can be evaluated both as a density model and as a policy-state classifier.

Finally, for scenario analysis, the gate can be frozen at a given value G . The implied long-run high-stress share of the resulting two-state transition matrix is

$$\omega_1(G) = \frac{p_{01}(G)}{p_{01}(G) + 1 - p_{11}(G)}, \quad p_{01}(G) = 1 - p_{00}(G). \quad (12)$$

This quantity converts transition coefficients into interpretable policy-state comparisons. For example, moving TNAC, TTF gas volatility, or the EUA activity proxy from a low to a high percentile changes $P_t(G)$ and therefore changes both the probability of entering stress and the persistence of stress once it begins. The empirical model is therefore not merely a volatility mixture; it is a regime process whose transition law is conditioned on observable features of the EU ETS market environment.

4 Estimation and forecast evaluation

The model is estimated on the policy-enriched common sample for which EUA returns and all active gate variables are observed jointly. Let

$$\mathcal{I} = \{t_0, t_0 + 1, \dots, t_{\max} - 1\}$$

denote this estimation index set, with sample size $n = |\mathcal{I}|$, beginning on May 15, 2019 in the empirical application. The timing convention is deliberately forward-looking: the standardized gate $G_t = (\tilde{s}_t, \tilde{\tau}_t, \tilde{g}_t, \tilde{v}_t)'$ is observed at time t and is used to determine the transition probabilities for Z_{t+1} . The realized return r_{t+1} is then evaluated under the predictive mixture distribution implied by the filtered probabilities and the policy-state gated transition matrix. This convention prevents the transition variables from being interpreted as ex post labels of the return state.

For a parameter vector

$$\theta = \{\mu_0, \mu_1, \sigma_0, \sigma_1, a_{00}, a_{10}, b_{00}, b_{10}\},$$

where

$$b_{i0} = \left(b_{i0}^{(s)}, b_{i0}^{(\tau)}, b_{i0}^{(g)}, b_{i0}^{(v)} \right)'$$

for the full gate and the appropriate subvectors are used for the nested specifications, the log likelihood is obtained from the Hamilton filter described in Section 3:

$$\ell(\theta) = \sum_{t \in \mathcal{I}} \log \left[\sum_{j=0}^1 \xi_{j,t+1|t}(\theta) \phi\{r_{t+1}; \mu_j, \sigma_j^2\} \right]. \quad (13)$$

The filtered probabilities are initialized with diffuse probabilities and updated recursively. The standard deviations are restricted to be positive, and the two regimes are labelled after estimation so that $\sigma_0 < \sigma_1$. This labelling makes state 0 the low-stress regime and state 1 the high-stress regime throughout the empirical analysis.

The likelihood is non-convex, so estimation uses a repeated-start design. All Markov-family specifications are estimated with the `statsmodels` BFGS optimizer using 20 search repetitions. The latent two-state benchmark uses a maximum of 1000 iterations, while the covariate-gated specifications use a maximum of 2000 iterations. The optimizer tolerance is the package default gradient tolerance, the reported covariance matrix is the inverse-Hessian-based covariance returned by the optimizer, and failed starts are handled by the repeated-search routine.

The estimation strategy uses maximum likelihood and compares a nested sequence of models on the same common sample. The first benchmark is a Gaussian iid model,

$$r_{t+1} = \mu + \sigma \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, 1),$$

which tests whether a single unconditional return distribution is sufficient. The second benchmark is a latent two-regime Markov normal model with constant transition probabilities. The other Markov specifications introduce observable gate variables progressively:

$$G_t^{(1)} = (\tilde{s}_t), \quad G_t^{(2)} = (\tilde{s}_t, \tilde{\tau}_t), \quad G_t^{(3)} = (\tilde{s}_t, \tilde{\tau}_t, \tilde{g}_t), \quad G_t^{(4)} = (\tilde{s}_t, \tilde{\tau}_t, \tilde{g}_t, \tilde{v}_t).$$

This nesting is central to the interpretation argument. A scarcity-only gate tests whether a price-based state proxy is sufficient. Adding TNAC tests whether official policy surplus carries incremental information. Adding TTF gas volatility tests the energy-stress channel. Adding the EUA activity proxy tests whether market participation affects regime entry or persistence after policy and energy variables have already been included.

The four baseline gate variables are pre-specified on economic grounds.

Reduced-form volatility benchmarks are included to separate two different notions of model quality. The Gaussian and Student- t GARCH(1,1) models are estimated as

$$r_{t+1} = \mu + \sqrt{h_{t+1}} z_{t+1}, \quad h_{t+1} = \omega + \alpha \varepsilon_t^2 + \beta h_t, \quad (14)$$

where z_{t+1} is either standard normal or standardized Student- t with degrees of freedom ν . In addition, the empirical comparison includes Student- t GJR-GARCH

and Student- t EGARCH specifications. These models are strong density benchmarks because they are designed to absorb volatility clustering, leverage-type asymmetry, and heavy tails. They are not direct substitutes for the gated Markov specification because they do not map regime probabilities to observable policy, energy, and activity variables.

All models are evaluated first by common-sample likelihood and information criteria,

$$\text{AIC} = -2\ell(\hat{\theta}) + 2k, \quad \text{BIC} = -2\ell(\hat{\theta}) + k \log n, \quad (15)$$

where k is the number of estimated parameters and n is the number of return observations. The relevant comparison is not only whether the proposed gate improves on volatility benchmarks in density fit, but whether it improves on the latent and nested Markov specifications as economically meaningful state variables are added. A gated model that fails this nested comparison would have weak interpretive content even if it generated plausible narratives.

Forecast and risk measures are computed from the one-step predictive distribution. For the policy-state gated Markov model, the reported quantities are the predictive mean m_{t+1} , predictive variance v_{t+1} , the 5 percent Value-at-Risk $q_{0.05,t+1}$, and the corresponding expected shortfall $ES_{0.05,t+1}$ from Equations (9)–(11). For the GARCH benchmarks, the same quantities are computed from the conditional innovation distribution and the fitted conditional variance. This gives a fair density and tail-risk comparison while preserving the distinction between reduced-form volatility forecasts and policy-state probabilities.

The second evaluation dimension is stress-state classification. Define absolute-tail and left-tail indicators by

$$I_t^{\text{abs}} = \mathbf{1}\{|r_t| \geq q_{0.90}(|r|)\}, \quad I_t^{\text{left}} = \mathbf{1}\{r_t \leq q_{0.05}(r)\}. \quad (16)$$

The gated Markov model is scored by the one-step-ahead predicted high-stress probability $\Pr(Z_t = 1 \mid \mathcal{F}_{t-1})$, whereas the volatility benchmarks are scored by their one-step-ahead conditional volatility. The reported area under the ROC curve (AUC) measures whether large realized stress days receive higher scores than ordinary days. The top-decile hit rate reports the fraction of tail days contained in the highest decile of each model score. In held-out validation, the absolute-tail and left-tail thresholds are fixed from the corresponding training sample to avoid future information in the event definition. To complement those ranking statistics, the analysis also evaluates Brier scores, predictive log scores, VaR hit rates, Kupiec unconditional-coverage tests, expected-shortfall comparisons, and block-bootstrap intervals for pairwise held-out score differences.

Finally, the fitted transition law is translated into policy-state scenarios. For each active driver, the driver is moved from its 10th to its 90th percentile while the other gate variables are fixed at their medians. The resulting transition probabilities $p_{01}(G)$ and $p_{11}(G)$, together with the implied high-stress share $\omega_1(G)$ in Equation (12), summarize the economic effect of TNAC, gas volatility, scarcity, and volume on regime entry and persistence. Event-window measures are used only as supporting evidence: for each policy or market event, the largest absolute return in a seven-day window

is recorded and compared with a conservative 2.5-standard-deviation threshold. The main empirical claim therefore rests on common-sample nested model comparison, state classification, and interpretable transition scenarios rather than on isolated event narratives.

5 Data and empirical design

The empirical analysis is built from a Refinitiv Workspace panel designed to separate three objects that are often mixed in carbon-market studies: the EUA return series to be modelled, the policy-state variables that can move the transition law, and supplementary market variables. The primary allowance series is the continuous EUA futures contract with RIC CFI2c1. The expanded market-state panel also contains Brent oil futures, WTI crude oil futures, Dutch TTF gas futures, German power futures, API2 coal, a European utilities index, Euro Stoxx 50, VSTOXX, and EUR/USD. Table 2 reports the Workspace identifiers, start dates, and observation counts. The longest price histories begin in early 2015 and run through May 30, 2026; the common estimation sample is shorter because all active transition variables must be jointly available and correctly timed with the next EUA return.

Let S_t denote the cleaned EUA futures price. Daily returns are computed as

$$r_t = 100\{\log(S_t) - \log(S_{t-1})\}, \quad (17)$$

so the reported regime means and standard deviations are in percent per trading day. The descriptive statistics use the longer aligned Refinitiv return span whenever the relevant price is available. The regime-switching comparison uses the policy-enriched common sample from May 15, 2019 to May 28, 2026, where EUA returns, the market-implied price-state proxy, official TNAC, Dutch TTF gas volatility, and EUA trading activity are observed jointly. This distinction is important: the long sample documents unconditional carbon-return properties, while the common sample is the valid sample for estimating the time-varying transition process in Section 3.

The active state vector is constructed to match the timing convention of the stochastic process. The gate $G_t = (\tilde{s}_t, \tilde{\tau}_t, \tilde{g}_t, \tilde{v}_t)'$ is observed at date t and is used to determine the transition distribution of Z_{t+1} before r_{t+1} is evaluated. In the implemented design, the target return is shifted forward by one trading day so that each row contains the gate observed at date t and the realized return from t to the next trading day. These four baseline gate variables are the pre-specified specification used for the main results. The price-state proxy s_t is a rolling percentile of the EUA futures price, interpreted as a market-implied relative price-state index rather than an external policy scarcity measure. For a rolling window of length L_s , it can be written as

$$s_t = \frac{1}{L_s} \sum_{\ell=1}^{L_s} \mathbf{1}\{S_{t-\ell} \leq S_t\}, \quad (18)$$

which places the current allowance price between 0 and 1 relative to its recent history. The empirical analysis sets $L_s = 252$ trading days. Official TNAC is measured in

Table 1 Construction of the active policy-state gate variables

Variable	Construction	Interpretation in the transition gate
Price-state proxy s_t	Rolling percentile of the EUA futures price	Market-implied relative EUA price state and banking value
TNAC τ_t	Official TNAC in millions of allowances, carried forward between publications	Policy surplus and MSR-relevant market balance
TTF gas volatility g_t	21-day annualized volatility of Dutch TTF gas returns	Energy-market stress and fuel-switching uncertainty
EUA log volume v_t	$\log(1 + V_t)$ from CFI2Zc1 ACVOL_UN\$	Trading intensity proxy; lagged and contract-matched alternatives are checked separately

millions of allowances and is carried forward only from its public release date onward, because TNAC is an annual policy-state variable rather than a daily market price (European Commission 2025, 2026b). Table 3 reports the publication-date audit used in the gate construction. Dutch TTF gas volatility is the 21-day annualized standard deviation of TTF log returns,

$$g_t = \sqrt{252} \text{sd}\{r_{t-20}^{\text{TTF}}, \dots, r_t^{\text{TTF}}\}, \quad (19)$$

and captures short-run energy-market stress. EUA trading activity is measured by log futures volume,

$$v_t = \log(1 + V_t), \quad (20)$$

where V_t is the observed EUA futures activity field. In the baseline specification, V_t is taken from Refinitiv RIC CFI2Zc1 field ACVOL_UN\$, which gives the broadest feasible common sample. Because that field is a December-contract activity proxy rather than a contract-matched continuous-return series, it is interpreted as a practical activity proxy rather than a contract-exact series. Each active regressor is standardized on the common estimation sample,

$$\tilde{x}_t = \frac{x_t - \bar{x}}{s_x}, \quad x_t \in \{s_t, \tau_t, g_t, v_t\}, \quad (21)$$

so that the transition coefficients in Section 6 are comparable across policy surplus, energy stress, and market activity.

Additional market and policy variables are retained as supplementary evidence rather than active baseline gates. German power volatility, API2 coal volatility, European utilities volatility, EUR/USD volatility, MSR threshold indicators, estimated MSR intake, and regulatory-event windows are used for robustness-oriented inspection only. The baseline gate is deliberately parsimonious: it contains one market-implied price-state variable, one official policy-state variable, one energy-stress variable, and

Table 2 Refinitiv Workspace download status for the expanded market-state panel

Series	RIC	Start	Observations
EUA continuous futures	CFI2c1	2015-01-02	2926
Brent oil futures	LC0c1	2015-01-02	2945
WTI crude oil futures	CLc1	2015-01-02	2872
Dutch TTF gas futures	TRNLTTFMc1	2015-01-02	2929
German power futures	DEBMc1	2017-04-24	2319
API2 coal	OTRAPI2Mc1~2	2015-01-02	2141
European utilities index	.SX6P	2015-01-02	2929
Euro Stoxx 50	.STOXX50E	2015-01-02	2928
VSTOXX	.V2TX	2015-01-02	2903
EUR/USD	EUR=	2015-01-02	2976

one market-activity variable. This design avoids an overfitted transition equation while preserving the economic channels required by the hypotheses.

The data-cleaning procedure is designed to make the timing convention explicit. First, raw Workspace series are sorted by date, converted to numeric values, and screened for non-positive prices before log returns are computed. Second, market-state variables are aligned to the EUA trading calendar; missing daily market variables are not mechanically forward-filled, while official policy variables are carried forward only from their public release dates. Third, the gate at date t is paired with the next trading day's return so that the transition law is genuinely forward looking at the row level. Fourth, the common estimation panel is formed by retaining only observations for which the forward return and all active gate variables are available. The event-window indicators are then added after the common sample is formed, so they support supplementary checks without defining the baseline transition mechanism. This procedure is intentionally conservative on volume handling: alternative activity fields are kept separate and compared explicitly rather than filled or merged ad hoc.

The resulting design links the empirical data directly to the econometric claim of the paper. The descriptive sample is used to show that EUA returns are heavy-tailed and volatile enough to require regime-sensitive modelling. The policy-enriched common sample is used to test whether observable market design, energy stress, and trading activity improve transition-state identification. The principal limitation is the low update frequency of official policy variables relative to daily returns; however, this is a feature of the EU ETS policy architecture rather than a data-processing defect. The empirical question is therefore whether this low-frequency official policy information, when combined with daily energy and trading variables, is sufficient to improve the interpretation of EUA return regimes.

6 Empirical results

The empirical results are organized around the economic logic of the model rather than around the order in which individual tables and figures are produced. The evidence is developed in four connected steps. First, the EUA return series is shown

Table 3 TNAC publication timing and first-use audit on the common sample.

Publication date	TNAC year	TNAC (million)	First gate date used	Last gate date used	Gate days in sample
2019-05-15	2018	1654.9098	2019-05-15	2020-05-07	249
2020-05-08	2019	1385.4962	2020-05-08	2021-05-14	259
2021-05-17	2020	1578.7724	2021-05-17	2022-05-12	255
2022-05-13	2021	1449.2142	2022-05-13	2023-05-12	256
2023-05-15	2022	1134.7947	2023-05-15	2024-05-31	267
2024-06-03	2023	1111.7365	2024-06-03	2025-05-30	252
2025-06-01	2024	1148.0496	2025-06-02	2026-05-28	251

to be too asymmetric, tail-sensitive, and clustered to be adequately summarized by a single Gaussian return law. Second, the transition gate is assessed as an interpretable mechanism linking policy scarcity, energy-market stress, and trading activity to movements between low- and high-stress regimes. Third, the fitted state probabilities are evaluated as stress-ranking and density objects. Fourth, the held-out evidence is used to discipline the strength of the forecasting claim. This structure separates three distinct claims: the presence of regime dependence, the economic interpretability of the transition variables, and the more demanding question of real-time predictive superiority.

6.1 From unconditional non-normality to latent stress regimes

The descriptive evidence in Table 4 provides the first reason for using a regime-sensitive specification. Over the aligned Refinitiv sample, the EUA proxy has an annualized mean log return of 21.59% and an annualized volatility of 42.81%. Its daily returns are mildly left-skewed, have excess kurtosis of 4.35, and produce a 5% expected shortfall of about -6.14% . These features indicate that the allowance return distribution contains both large downside observations and non-Gaussian tails. The comparison with energy and financial variables reinforces this interpretation. Dutch TTF gas is substantially more volatile, with annualized volatility of 74.41% and a 5% expected shortfall of about -10.73% , while Brent, WTI, utilities, Euro Stoxx 50, volatility, and EUR/USD series show that EUA futures are embedded in a broader energy-financial system. The basic empirical environment is therefore not a one-factor carbon market in isolation, but a policy-created asset exposed to scarcity, energy shocks, macro-financial risk, and trading conditions.

The time-series evidence in Figure 1 shows that this non-normality is dynamic rather than purely marginal. EUA prices rose strongly after 2018, were sharply repriced during the 2021–2022 energy crisis, corrected during the later allowance-price draw-down, and then experienced further recovery and adjustment episodes. Large positive and negative returns are temporally clustered, and the filtered and one-step-ahead predicted high-stress probabilities concentrate around the 2022 energy shock, the 2024 allowance-price drawdown and rebound, and the early-2026 adjustment. The filtered probability describes the state associated with the return just observed, whereas the

Table 4 Summary statistics for Refinitiv market series.

Series	Obs.	Ann. mean	Ann. vol.	Skew	Ex. kurt.	ES 5%
EUA proxy	2896	21.5916	42.8108	-0.3063	4.3504	-6.1444
Brent oil	2909	4.6007	40.3404	-1.1848	17.2974	-6.0516
Crude oil	2763	18.5271	46.5316	0.6981	34.2154	-6.3534
Dutch TTF gas	2858	1.9858	74.4108	0.7361	12.7036	-10.7310
API2 coal	612	-12.4832	35.1932	0.4418	11.9892	-5.0623
European utilities	2897	4.2596	17.5224	-1.4247	16.9675	-2.6281
EuroStoxx50	2895	5.8514	18.9186	-0.8006	10.7326	-2.9145
VSTOXX	2859	-16.0713	113.8845	0.7229	4.2616	-14.3719
EURUSD	2975	-0.2442	7.7844	0.1377	2.2592	-1.0678

predicted probability is the timing-correct score used in the ranking exercises below. Thus, the state sequence is interpreted as a compact representation of clustered market stress rather than as independent evidence of forecasting dominance.

Distributional evidence in Figure 2 leads to the same conclusion from a different angle. The empirical histogram has a sharper center and fatter tails than a Gaussian density, and the Q-Q plot shows visible tail departures from normality, especially on the downside. A two-regime mixture provides a parsimonious way to combine ordinary low-stress observations with a high-volatility component. This link between clustered stress probabilities and fat-tailed returns is the main support for H1: the Gaussian iid benchmark is not an adequate description of EUA futures returns, and regime dependence is an empirically relevant feature of the data.

6.2 Information added by the policy-state transition gate

Once regime dependence is established, the next question is whether observable policy and market variables improve the transition law. The model comparison in Table 5 gives a clear within-family answer but also imposes an important boundary on the interpretation. Moving from the Gaussian iid model to the latent two-regime Markov normal model lowers the AIC from 8442.40 to 8198.23, confirming that regime dependence is essential. The price-state-only gate does not improve the latent benchmark on the common sample, with an AIC of 8201.58. The improvement begins when policy and market-state information is added: including TNAC lowers the AIC to 8192.02, adding Dutch TTF gas volatility reduces it to 8182.67, and the full specification with the EUA activity proxy reaches 8176.34. This nested progression supports the transition-gate argument within the Gaussian Markov class.

The same evidence also prevents overstatement. The full gate is selected by log likelihood and AIC, but not by BIC. Its BIC is 8253.19, compared with 8231.17 for the latent Markov normal benchmark and 8248.55 for the policy-plus-gas gate. Heavy-tailed GARCH-family models are stronger unconditional density engines, with AIC/BIC pairs of 8116.46/8143.90 for Student- t GARCH and 8116.35/8149.29 for GJR-GARCH Student- t . The appropriate claim is therefore not that the full gate dominates all volatility models. Rather, observable policy, energy, and activity variables add interpretable transition information within a Gaussian regime-switching



Fig. 1 Refinitiv EUA futures price, returns, and estimated high-stress probability. The shaded area in the lower panel is the filtered high-stress probability, which uses the realized return at date t . The line is the one-step predicted high-stress probability, which conditions only on $\mathcal{F}_{t-1}$ and is the score used in the ranking analysis.

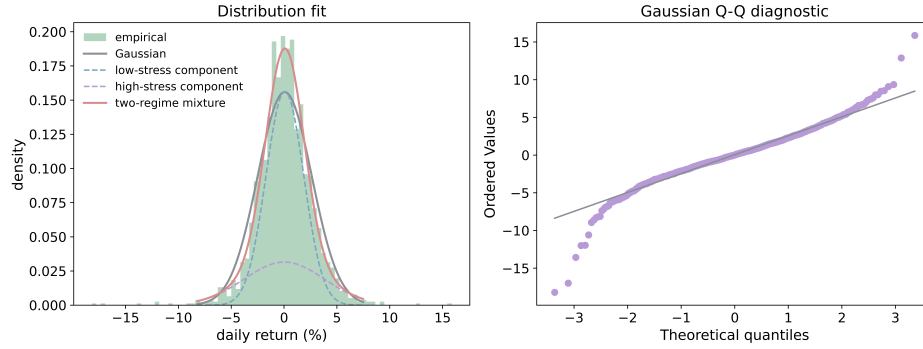


Fig. 2 Distributional evidence for Refinitiv EUA futures returns.

Table 5 Common-sample model comparison for EUA futures returns.

Model	Log likelihood	AIC	BIC
Gaussian iid	-4219.2019	8442.4038	8453.3826
Latent two-regime Markov normal	-4093.1171	8198.2342	8231.1707
Scarcity-only Markov gate	-4092.7905	8201.5810	8245.4963
Scarcity + TNAC Markov gate	-4086.0097	8192.0193	8246.9134
Policy + gas-vol Markov gate	-4079.3366	8182.6731	8248.5461
Proposed policy-gated Markov normal	-4074.1707	8176.3413	8253.1931
GARCH(1,1) normal	-4081.2826	8170.5651	8192.5228
GARCH(1,1) Student-t	-4053.2289	8116.4577	8143.9048
GJR-GARCH(1,1) Student-t	-4052.1771	8116.3541	8149.2906
EGARCH(1,1) Student-t	-4053.2855	8118.5710	8151.5075

representation, while heavy-tailed conditional-volatility models remain harder density benchmarks.

The economic role of the gate is clarified by the construction of the state variables in Figure 3. The price-state proxy summarizes relative EUA price pressure, TNAC and MSR intake represent slowly updated official surplus information, Dutch TTF gas volatility captures the dominant energy-stress channel, and the Refinitiv activity proxy captures changes in allowance-market trading intensity. These variables are not redundant controls. They represent different channels through which a policy-created commodity-financial asset can enter, remain in, or exit a high-stress state.

At the parameter level, Tables 6 and 7 show that the two regimes are separated primarily by risk rather than by expected return. The lower-stress state has an estimated daily mean of about $\mu_0 = 0.077\%$ and a daily standard deviation of about $\sigma_0 = 1.788\%$. The higher-stress state has a similar mean, about $\mu_1 = 0.024\%$, but a much larger daily standard deviation of about $\sigma_1 = 3.790\%$. This is the pattern expected from a stress-state model: the high regime is not a high-return state, but a high-uncertainty state. The benchmark volatility coefficients are reported separately so that the interpretation of Markov transition risk is not conflated with reduced-form GARCH variance dynamics.

The transition coefficients provide the main associative interpretation of the gate. Because the logistic transition law is written for $\Pr(Z_{t+1} = 0 \mid Z_t = i, \cdot)$, a negative slope in the $0 \rightarrow 0$ equation lowers low-state persistence and raises the probability of entering stress. TNAC and gas volatility both have negative $0 \rightarrow 0$ slopes, so higher values are associated with greater pressure to move from the low-stress state to the high-stress state. The activity proxy behaves differently. Its $0 \rightarrow 0$ slope is positive, while its $1 \rightarrow 0$ slope is negative, which is more consistent with a stress-persistence channel than with a simple stress-entry trigger. The price-state proxy is also asymmetric: a higher relative EUA price level is associated with more stress-entry pressure through the $0 \rightarrow 0$ equation, but faster exit from stress through the $1 \rightarrow 0$ equation.

Scenario calculations in Table 8 make these coefficient signs economically readable by moving each driver from its 10th to its 90th percentile while holding the remaining drivers at their medians. TNAC is the clearest policy-state channel in point-estimate

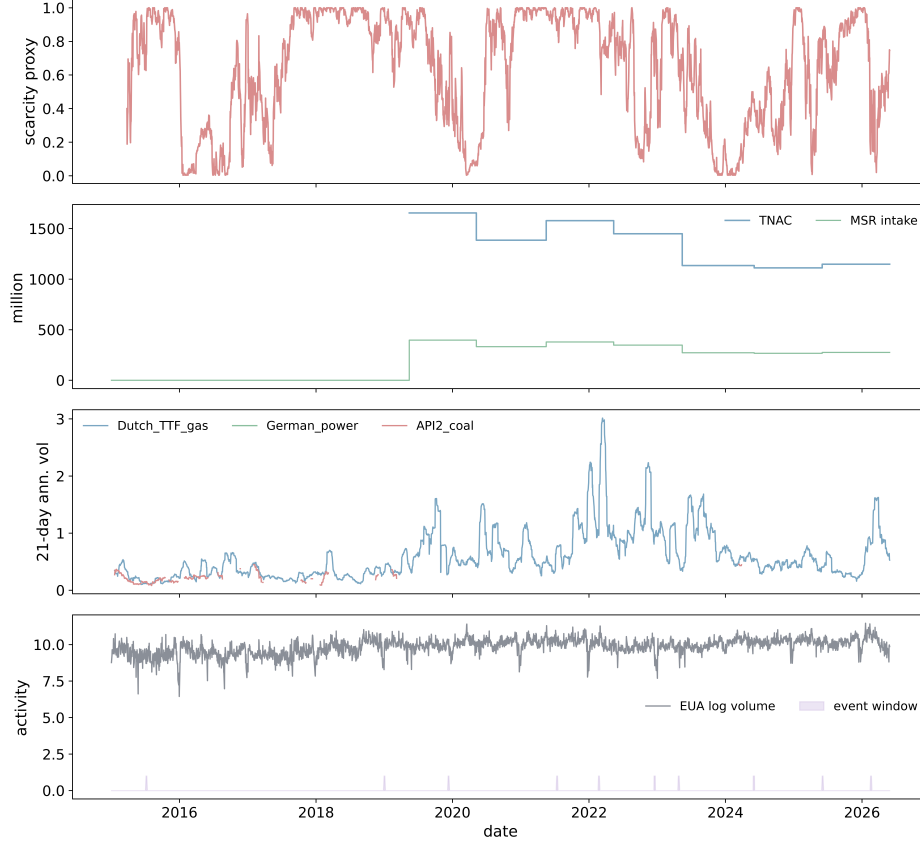


Fig. 3 Observable policy-state variables used in the baseline design. The scarcity panel is a unit-interval rolling price-state proxy. The TNAC and MSR panel is in millions of allowances. The energy panel reports 21-day annualized volatilities. The activity panel reports $\log(1 + V_t)$ from Refinitiv RIC CFI2Zc1 field ACVOL.UNS, with shaded regulatory-event windows shown as supplementary context.

terms: when TNAC rises from 1,111.7 to 1,654.9 million allowances, the low-to-high transition probability increases from 0.83% to 21.05%, and the implied stress-state share rises from 7.76% to 31.60%. This is economically meaningful, but it must be interpreted cautiously because TNAC changes only at annual publication dates. The gas-volatility channel is stronger and more precisely estimated. When Dutch TTF gas volatility rises from 0.34 to 1.39, the low-to-high transition probability increases from 2.07% to 25.20%, and the implied stress-state share rises from 8.68% to 47.69%. The activity proxy again points to persistence rather than entry: moving it from 9.64 to 10.56 lowers the low-to-high transition probability from 7.59% to 2.78%, but raises high-stress persistence enough to increase the implied stress-state share from 11.24% to 30.69%. Finally, the market-implied price-state proxy lowers the implied stress share from 26.16% to 15.29% across its 10th-to-90th percentile shift, confirming that its role is asymmetric rather than a monotone scarcity effect.

Table 6 Estimated parameters of the covariate-gated two-regime Markov normal model. Entries report estimates with delta-method standard errors in brackets.

Symbol	Meaning	Estimate [s.e.]
a_{00}	transition intercept, $0 \rightarrow 0$	3.064 [0.361]
a_{10}	transition intercept, $1 \rightarrow 0$	-1.528 [0.507]
$b_{00}^{(s)}$	scarcity slope, $0 \rightarrow 0$	-0.725 [0.342]
$b_{10}^{(s)}$	scarcity slope, $1 \rightarrow 0$	1.165 [0.431]
$b_{00}^{(\tau)}$	TNAC slope, $0 \rightarrow 0$	-1.316 [0.339]
$b_{10}^{(\tau)}$	TNAC slope, $1 \rightarrow 0$	0.773 [0.520]
$b_{00}^{(g)}$	gas-volatility slope, $0 \rightarrow 0$	-1.222 [0.385]
$b_{10}^{(g)}$	gas-volatility slope, $1 \rightarrow 0$	0.142 [0.347]
$b_{00}^{(v)}$	volume slope, $0 \rightarrow 0$	0.452 [0.250]
$b_{10}^{(v)}$	volume slope, $1 \rightarrow 0$	-1.332 [0.469]
μ_0	low-stress mean, percent	0.077 [0.059]
μ_1	high-stress mean, percent	0.023 [0.178]
σ_0	low-stress s.d., percent	1.788 [0.054]
σ_1	high-stress s.d., percent	3.790 [0.176]

Table 7 Estimated parameters of benchmark GARCH-family models.

Symbol	Meaning	GARCH(1,1) normal	GARCH(1,1) Student- t	GJR-GARCH Student- t	EGARCH Student- t
μ	constant mean, percent	0.062	0.070	0.0580	0.0570
ω	variance intercept	0.126	0.107	0.0930	0.0470
α_1	shock loading	0.091	0.074	0.0530	0.1660
γ_1	asymmetry loading	—	—	0.0280	-0.0220
β_1	persistence loading	0.893	0.911	0.9200	0.9760
ν	Student- t d.f.	—	7.684	7.6990	7.3140

The uncertainty quantification in Table 9 narrows these scenario interpretations. The gas-volatility effect remains the strongest channel, with a stress-share increase of 0.3901 and a confidence interval from 0.1323 to 0.6336. The TNAC point estimate is positive, but its stress-share interval crosses zero, which is consistent with the low effective number of official policy updates. The activity effect is also delicate: its point estimate raises the stress share through persistence, but both the stress-share and low-to-high-probability intervals are wide. Consequently, H2 is supported only cautiously, mainly through the AIC improvement and the economically meaningful TNAC point estimate, while H3 is supported most clearly for the gas-volatility channel and more tentatively for the activity channel.

Table 8 Implied transition probabilities across policy-state scenarios.

Driver	Level	Value	$\Pr(Z_{t+1} = 1 Z_t = 0)$	$\Pr(Z_{t+1} = 1 Z_t = 1)$	Implied stress share
Market-implied price-state proxy	10th percentile	0.1306	0.0129	0.9634	0.2608
Market-implied price-state proxy	90th percentile	0.9881	0.0827	0.5421	0.1530
TNAC (million allowances)	10th percentile	1111.7365	0.0083	0.9013	0.0774
TNAC (million allowances)	90th percentile	1654.9098	0.2104	0.5447	0.3160
TTF gas volatility	10th percentile	0.3433	0.0206	0.7829	0.0867
TTF gas volatility	90th percentile	1.3852	0.2524	0.7234	0.4771
EUA activity proxy	10th percentile	9.6379	0.0759	0.4007	0.1124
EUA activity proxy	90th percentile	10.5626	0.0278	0.9373	0.3069

Table 9 Simulation-based uncertainty intervals for scenario-implied transition effects.

Driver	10th pct.	90th pct.	$\Delta \Pr(Z_{t+1} = 1 Z_t = 0)$	CI L	CI H	Δ stress share	CI L	CI H
Market-implied price-state proxy	0.1306	0.9881	0.0697	0.0096	0.1807	-0.1079	-0.4543	0.1425
TNAC (million allowances)	1111.7365	1654.9098	0.2021	0.0810	0.4057	0.2386	-0.0284	0.4457
TTF gas volatility	0.3433	1.3852	0.2318	0.0619	0.5137	0.3904	0.1327	0.6336
EUA activity proxy	9.6379	10.5626	-0.0481	-0.1138	0.0057	0.1945	-0.0453	0.5022

6.3 Stress ranking, predictive density, and timing discipline

The one-step-ahead high-stress probability can be used as a risk-ranking score, but the timing convention is crucial. Table 10 reports this comparison with stress labels defined from realized returns at date t while the Markov score conditions only on information available at $t - 1$. Under the common-sample exercise, with parameters estimated on the full common sample, the Markov score obtains an AUC of 0.6993 for absolute-tail days and 0.6893 for 5% left-tail days. The Gaussian, Student- t , GJR-GARCH, and EGARCH benchmarks lie between 0.6650 and 0.6717 for absolute-tail days and between 0.6488 and 0.6564 for left-tail days. The gain is therefore modest. The top score decile captures 23.46% of absolute-tail observations under the Markov score, compared with about 21–22% under the volatility benchmarks, while left-tail top-decile capture is not superior. These results indicate that the transition score improves the ranking of some full-sample stress observations, although it is not a uniformly dominant stress classifier.

The fitted predictive density reinforces this interpretation. Table 11 summarizes the implied one-day distribution, and Table 12 compares it with the volatility benchmarks. In percent return units, the one-day predictive mean is approximately 0.075%, the predictive standard deviation is 1.819%, the 5% Value-at-Risk is about -2.89% , and the 5% expected shortfall is about -3.71% . The probability of a positive return

Table 10 Common-sample one-step-ahead ranking results using predicted regime probabilities and volatility benchmarks.

Model	AUC: abs-tail days	AUC: left-tail days	Top-decile abs-tail share (%)	Top-decile left-tail share (%)
Policy-gated Markov	0.6993	0.6893	23.4637	10.0559
GARCH normal	0.6699	0.6564	21.2291	11.7318
GARCH Student- t	0.6717	0.6512	21.2291	11.7318
GJR-GARCH Student- t	0.6717	0.6503	21.2291	11.1732
EGARCH Student- t	0.6650	0.6488	21.7877	11.1732

Table 11 Predictive distribution implied by the stochastic process.

Quantity	Value
Next-day mean	0.0755
Next-day standard deviation	1.8193
5 percent Value-at-Risk	-2.8850
5 percent expected shortfall	-3.7093
Probability of a positive return	0.5170

Table 12 Predictive density and tail-risk comparison across regime and volatility models.

Model	Brier LT	Log score LT	Brier AT	VaR 5% hit	Kupiec p	Mean exceed.	Mean ES
Policy-gated Markov	0.0470	0.1896	0.0862	4.3041	0.1669	-5.6818	-5.2755
GARCH normal	0.0479	0.1950	0.0890	4.4159	0.2478	-5.6247	-4.9200
GARCH Student- t	0.0476	0.1941	0.0879	4.8072	0.7065	-5.5062	-5.1987
GJR-GARCH Student- t	0.0477	0.1945	0.0879	4.6954	0.5505	-5.5782	-5.2480
EGARCH Student- t	0.0474	0.1942	0.0871	4.5836	0.4127	-5.6893	-5.3361

is 0.517, close to one half, so the predictive distribution is not driven by a large directional mean. Its risk content comes mainly from the mixture of low- and high-stress conditional variances. On the full common sample, the Markov model has slightly better Brier and predictive log scores for left-tail events and the best absolute-tail Brier score. Its full-sample 5% VaR hit rate is 4.30%, with a Kupiec p -value of 0.167, indicating acceptable common-sample unconditional coverage. Nevertheless, the heavy-tailed GARCH benchmarks remain stronger by likelihood and information criteria, which keeps the density claim conservative.

The held-out validation exercise gives a stricter assessment of forecasting performance. Tables 13 and 14 report the split-based results using a pre-crisis training sample through December 2020, a crisis validation sample for 2021–2022, and a post-crisis test sample from 2023 onward, with tail-day thresholds fixed from the relevant

Table 13 Held-out ranking and scoring performance under fixed train-validation-test splits. Tail-day labels are defined with thresholds fixed from the corresponding training sample. Model labels are MS-Markov, GARCH-N, GARCH- t , GJR- t , and EGARCH- t .

Segment	Model	Obs.	Avg. log score	AUC AT	AUC LT	Top-decile AT (%)	Top-decile LT (%)
Crisis val.	MS-Markov	511	-2.5084	0.4818	0.5084	13.4615	9.6154
Crisis val.	GARCH-N	511	-2.4518	0.5884	0.6017	17.3077	7.6923
Crisis val.	GARCH- t	511	-2.4265	0.5882	0.6011	17.3077	9.6154
Crisis val.	GJR- t	511	-2.4302	0.5620	0.5718	19.2308	7.6923
Crisis val.	EGARCH- t	511	-2.4312	0.5520	0.5456	21.1538	11.5385
Post-crisis	MS-Markov	862	-2.2681	0.5744	0.4880	8.0460	3.4483
Post-crisis	GARCH-N	862	-2.1498	0.7499	0.7018	5.7471	2.2989
Post-crisis	GARCH- t	862	-2.1356	0.7662	0.7137	5.7471	2.2989
Post-crisis	GJR- t	862	-2.1425	0.7506	0.6807	4.5977	2.2989
Post-crisis	EGARCH- t	862	-2.1454	0.7617	0.6685	6.8966	2.2989

Table 14 Held-out tail-risk calibration under fixed train-validation-test splits. Tail-day labels are defined with thresholds fixed from the corresponding training sample. Model labels are MS-Markov, GARCH-N, GARCH- t , GJR- t , and EGARCH- t .

Segment	Model	Brier LT	Brier AT	VaR 5% hit	Kupiec p	Mean exceed.	Mean ES
Crisis val.	MS-Markov	0.0609	0.0960	5.8708	0.3788	-6.7206	-5.9471
Crisis val.	GARCH-N	0.0612	0.0940	6.0665	0.2837	-6.6734	-5.7385
Crisis val.	GARCH- t	0.0608	0.0928	6.0665	0.2837	-6.6734	-6.2263
Crisis val.	GJR- t	0.0611	0.0933	5.2838	0.7705	-6.8760	-6.3365
Crisis val.	EGARCH- t	0.0610	0.0926	5.6751	0.4926	-6.7850	-6.3303
Post-crisis	MS-Markov	0.0207	0.0498	0.9281	0.0000	-5.8208	-8.5593
Post-crisis	GARCH-N	0.0151	0.0312	2.4362	0.0001	-4.9033	-4.9627
Post-crisis	GARCH- t	0.0151	0.0311	2.3202	0.0001	-4.9806	-5.4731
Post-crisis	GJR- t	0.0153	0.0318	2.2042	0.0000	-5.1688	-5.6066
Post-crisis	EGARCH- t	0.0155	0.0323	1.8561	0.0000	-5.3661	-5.8121

training sample. In these held-out segments, the transition model does not dominate the heavy-tailed volatility benchmarks. During the crisis validation period, its AUC is 0.464 for absolute-tail days and 0.518 for left-tail days. In the post-crisis test period, it is weaker on log score, Brier score, and AUC, and its 5% VaR hit rate is 0.93%.

The block-bootstrap comparisons in Table 15 further qualify the forecasting interpretation. Relative to Student- t GARCH and GJR-GARCH Student- t , the Markov model has lower average log scores in both held-out segments, with bootstrap intervals for Δ average log score entirely below zero. In the post-crisis test segment, it also has higher Brier scores and lower absolute-tail AUC, with intervals excluding zero. Left-tail AUC differences are less precisely determined, but the practical conclusion is unchanged: the model has interpretive value for policy-state regime analysis, whereas the held-out evidence does not support a stronger claim of superior real-time stress prediction.

Table 15 Block-bootstrap uncertainty intervals for held-out forecast-comparison differences. Positive Δ AUC favors the Markov model.

Segment	Benchmark	Δ log score	CI L	CI H	Δ Brier LT	CI L	CI H	Δ Brier AT	CI L	CI H	Δ AUC AT
Crisis val.	GARCH- t	-0.0807	-0.1863	-0.0180	0.0002	-0.0011	0.0016	0.0037	-0.0005	0.0101	-0.1198
Crisis val.	GJR- t	-0.0771	-0.1884	-0.0162	-0.0001	-0.0018	0.0016	0.0033	-0.0012	0.0095	-0.0935
Post-crisis	GARCH- t	-0.1324	-0.1592	-0.1056	0.0056	0.0046	0.0064	0.0186	0.0157	0.0216	-0.1917
Post-crisis	GJR- t	-0.1255	-0.1495	-0.0994	0.0054	0.0043	0.0063	0.0180	0.0149	0.0205	-0.1761

6.4 Synthesis of the hypotheses

Overall, the results support the hypotheses with different degrees of strength. H1 is strongly supported because EUA returns are fat-tailed, downside-sensitive, and clustered, and because the latent two-regime model materially improves on the Gaussian iid benchmark. H2 receives cautious support: TNAC improves the nested Markov specification and has economically meaningful scenario effects, but its low update frequency produces wide uncertainty intervals. H3 is supported most clearly through the gas-volatility channel, which has the strongest and most precisely estimated effect on the implied stress-state share; the activity channel is better interpreted as a stress-persistence proxy and is less precisely identified. H4 is supported as a common-sample regime-ranking result. The one-step transition score ranks some common-sample stress days better than volatility-based scores, but the held-out validation and bootstrap comparisons do not support a stronger real-time forecasting claim.

The empirical contribution is therefore deliberately bounded. The proposed policy-state gated Markov model gives an economically interpretable map from policy surplus, energy stress, and trading activity to the probability of being in, entering, or remaining in a high-stress EUA return regime. The specification is not intended as a causal estimate of regulatory effects or as a universal density-forecasting model. Its value is to connect allowance-market risk to observable policy and market-state variables in a way that is transparent, timing-aware, and compatible with risk-management applications.

7 Conclusion and implications

This paper develops a time-varying transition probability return-regime process for a market whose traded asset is created by regulation. The central argument is that EUA futures returns should not be modelled only as a generic financial time series. Their distribution is shaped by the institutional design of the EU ETS, the surplus-management role of the MSR, short-run energy-market pressure, and the trading activity of compliance and financial participants. The proposed specification therefore places observable policy, energy, and market-activity variables directly inside the Markov transition law, rather than attaching those variables only as ex post explanations for latent volatility states.

The empirical results support this interpretation with a bounded scope. The descriptive evidence shows that EUA futures returns are heavy-tailed, asymmetric in the tails, and materially regime dependent, so a single Gaussian return law is not an adequate representation of the data. Within the Markov family, the nested model

comparison shows that a purely latent benchmark is improved when the transition gate contains policy and market information. A price-state proxy alone is insufficient, whereas the addition of official TNAC, Dutch TTF gas volatility, and an activity proxy produces a more informative transition specification. This confirms that the relevant object is not merely regime switching, but regime switching whose transition probabilities can be related to observable EU ETS conditions.

The fitted transition specification also gives economically interpretable conclusions. Higher TNAC and higher TTF gas volatility are associated with a higher probability that the process occupies the high-stress state, although the TNAC effect is estimated from a small number of annual step changes and is therefore less precise than the market-variable effects. The activity proxy behaves mainly as a persistence channel, and its quantitative strength depends on the Refinitiv field used. These results convert coefficient estimates into scenario-based changes in transition probabilities and stress shares that are directly relevant to carbon-market researchers, risk managers, and policy analysts.

The comparison with GARCH clarifies the scope of the contribution. Heavy-tailed GARCH models remain stronger unconditional density benchmarks on the common sample, and they also outperform the transition model in the held-out validation and post-crisis test segments. The Markov model contributes on a different dimension. It provides a transition specification that can be interpreted in terms of TNAC, gas-market volatility, and EUA market activity, and it delivers a state-probability summary that is useful for common-sample regime interpretation. Hence the contribution is associative and interpretive rather than a claim of causal identification or universal forecasting dominance.

Several limitations also point to productive extensions of the present framework. Official policy-state variables such as TNAC are observed at much lower frequency than daily futures returns, so the estimated gate is best interpreted as a persistent state specification rather than a high-frequency causal shock model. The baseline specification also leaves room to incorporate additional information on auction calendars, allowance banking, verified emissions, cap-path revisions, options-implied uncertainty, and more granular MSR implementation details. In addition, the evidence in this paper is developed within a two-state Gaussian Markov family, so Markov- t , skewed Markov, or MS-GARCH extensions provide natural directions for broader density benchmarking. The held-out results likewise suggest that the transition model is more useful for economic interpretation and regime analysis than for stand-alone real-time risk forecasting. With these qualifications, the empirical evidence supports a focused conclusion: policy surplus, energy stress, and trading activity help interpret the law of motion for emission allowance return regimes.

Acknowledgements. This work was supported by Thammasat University under the Fundamental Fund (FF) scheme, Fiscal Year 2026, through the research project entitled “Integrating Carbon Credits into Portfolio Strategies for Risk Reduction from Carbon Emission Limitation Policies.”

Declarations

Funding. This work was supported by Thammasat University under the Fundamental Fund (FF) scheme, Fiscal Year 2026, through the research project entitled “Integrating Carbon Credits into Portfolio Strategies for Risk Reduction from Carbon Emission Limitation Policies.”

Competing interests. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability. Data will be made available on request.

Author contributions. Wanyok Attisattapong: Resources, Methodology, Investigation, Project administration, Supervision, Funding acquisition, and Conceptualization. Pasin Marupanthorn: Writing – original draft, Investigation, Validation, Resources, Formal Analysis, Data curation, Quantitative Modeling, and Conceptualization.

Ethics approval and consent to participate. Not applicable.

Clinical trial number. Not applicable.

Consent for publication. Not applicable.

Materials availability. Not applicable.

Code availability. Code will be made available on reasonable request.

References

- Bai, Y., Okullo, S.J.: Drivers and pass-through of the eu ets price: Evidence from the power sector. *Energy Economics* **123**, 106698 (2023) <https://doi.org/10.1016/j.eneco.2023.106698>
- Borghesi, S., Pahle, M., Perino, G., Quemin, S., Willner, M.: The market stability reserve in the eu emissions trading system: A critical review. *Annual Review of Resource Economics* **15**, 131–152 (2023)
- Benz, E., Trueck, S.: Modeling the price dynamics of co2 emission allowances. *Energy Economics* **31**(1), 4–15 (2009)
- Cepni, O., Gil-Alana, L.A., Gupta, R., Polat, O.: Time-variation in the persistence of carbon price uncertainty: The role of carbon policy uncertainty. *The Quarterly Review of Economics and Finance* **102**, 102004 (2025) <https://doi.org/10.1016/j.qref.2025.102004>
- Demiralay, S., Gencer, H.G., Bayraci, S.: Carbon credit futures as an emerging asset: Hedging, diversification and downside risks. *Energy Economics* **113**, 106196 (2022)
- Daskalakis, G., Psychoyios, D., Markellos, R.N.: Modeling co2 emission allowance prices and derivatives: Evidence from the european trading scheme. *Journal of Banking and Finance* **33**(7), 1230–1241 (2009)
- European Parliament and Council of the European Union: Directive (EU) 2023/959 of the European Parliament and of the Council of 10 May 2023 Amending Directive

- 2003/87/EC Establishing a System for Greenhouse Gas Emission Allowance Trading within the Union and Decision (EU) 2015/1814 Concerning the Market Stability Reserve (2023). <https://eur-lex.europa.eu/eli/dir/2023/959/oj/eng>
- European Commission: Publication of the Total Number of Allowances in Circulation in 2024 for the Purposes of the Market Stability Reserve under the EU Emissions Trading System (2025). https://climate.ec.europa.eu/document/download/afa7e5cf-4168-4a63-bdad-627db770b3bf_en
- European Commission: About the EU Emissions Trading System (2026). https://climate.ec.europa.eu/eu-action/carbon-markets/about-eu-ets_en
- European Commission: Market Stability Reserve (2026). https://climate.ec.europa.eu/eu-action/carbon-markets/eu-emissions-trading-system-eu-ets/market-stability-reserve_en
- Eksi, Z., Yerli, C., Arıkan, E.: Modeling and Estimation of Carbon Allowance Prices with Regime-Switching Dynamics in a Partial Information Setting. SSRN working paper (2025). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5336963
- Filardo, A.J.: Business-cycle phases and their transitional dynamics. *Journal of Business & Economic Statistics* **12**(3), 299–308 (1994) <https://doi.org/10.1080/07350015.1994.10524545>
- Fuchs, M., Stroebel, J., Terstege, J.: Carbon VIX: Carbon Price Uncertainty and Decarbonization Investments. NBER working paper (2024). <https://www.nber.org/papers/w32937>
- Gronwald, M., Ketterer, J.: What Moves the European Carbon Market? Insights from Conditional Jump Models. CESifo Working Paper No. 3795 (2012). https://www.ifo.de/DocDL/cesifo1_wp3795.pdf
- Hamilton, J.D.: A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* **57**(2), 357–384 (1989) <https://doi.org/10.2307/1912559>
- Li, C.-Y., Chen, S.-N., Lin, S.-K.: Pricing derivatives with modeling co2 emission allowance using a regime-switching jump diffusion model. *The European Journal of Finance* **22**(10), 887–908 (2016)
- Pahle, M., Quemin, S., Osorio, S., Günther, C., Pietzcker, R.: The emerging endgame: The eu ets on the road towards climate neutrality. *Resource and Energy Economics* **81**, 101476 (2025) <https://doi.org/10.1016/j.reseneeco.2024.101476>
- Perino, G., Willner, M., Quemin, S., Pahle, M.: The european union emissions trading system market stability reserve: Does it stabilize or destabilize the market? *Review of Environmental Economics and Policy* **16**(2), 338–345 (2022) <https://doi.org/10.1016/j.reneeco.2022.101476>

- Seifert, J., Uhrig-Homburg, M., Wagner, M.: Dynamic behavior of co2 spot prices. *Journal of Environmental Economics and Management* **56**(2), 180–194 (2008)
- World Bank Group: State and Trends of Carbon Pricing 2026 (2026). <https://www.worldbank.org/en/publication/state-and-trends-of-carbon-pricing>
- Wei, P., Zhou, J., Ren, X., Huynh, L.D.T.: Financialisation of the european union emissions trading system and its influencing factors in quantiles. *International Journal of Finance & Economics* **30**(1), 925–940 (2025) <https://doi.org/10.1002/ijfe.2950>