

Model Risk in European Union Allowance Futures and Options: From Black–76 to Institutionally Grounded Models—A Critical Structured Review

Bohao Zhu*

Beijing Jiaotong University, Weihai, China

**Corresponding author: Bohao Zhu.*

Abstract

Pricing models for European Union Allowance (EUA) derivatives are often assessed as statistical specifications even though the EU Emissions Trading System changes the payoff boundary, supply process, and intertemporal value of the underlying compliance asset. This critical structured review examines how EUA futures and options models have developed from Black–76 conventions to institutionally grounded equilibrium, structural, and Market Stability Reserve (MSR) specifications; how regulatory mechanisms enter risk-neutral valuation; and what evidence supports model robustness. Fifty-three DOI- or publisher-verified academic records were coded using an Institution–Model–Measure–Evidence (IMME) lens: 47 inform the model-level synthesis and six provide contextual triangulation. The evidence is organized into early benchmarks and deficiencies, institutionally grounded models, and MSR-era pricing and validation. Black–76 remains a transparent quotation benchmark, but evidence from physical-measure dynamics and phase boundaries limits its structural interpretation. Structural and equilibrium models make institutional channels explicit, yet add identification and numerical risks. Comparative studies provide some held-out pricing evidence, but the literature still lacks transaction-cost-aware hedging and regime-transport tests. The resulting model-selection principle is conditional adequacy: use the least complex model that is regime-aligned, measure-explicit, empirically identified, and validated for the intended pricing or hedging task.

Keywords

EU emissions trading system, EUA futures and options, model risk, risk-neutral valuation, market stability reserve

1. Introduction

European Union Allowances (EUAs) are both tradable assets and regulator-created compliance instruments. Their futures and options are quoted in a language inherited from conventional commodity derivatives, and Black–76 remains the natural reference for a European option written on a futures contract [1]. Yet a transparent benchmark is not the same as an adequate model. Model risk arises when several statistically plausible specifications imply materially different prices, sensitivities, reserves, or hedge outcomes [2]. In the

EUA market, this risk is unusually consequential because the stochastic process is not determined only by market information: it also reflects rules that can change the survival, scarcity, and terminal value of an allowance.

The institutional architecture makes cross-period transport especially difficult. Banking and borrowing determine whether scarcity can be shifted through time [3]; compliance calendars, penalties, allocation, and verified-emissions disclosure affect the payoff boundary and information set; and reforms such as the Market Stability Reserve (MSR) make effective supply depend on the state of the market [4–6]. Consequently, a model may be arbitrage-free and fit one period while remaining economically misspecified for another phase. In-sample accuracy can also conceal a misspecified transition from the physical probability measure to the risk-neutral measure, weak identification from sparse option quotes, or a hedge that fails when liquidity and policy conditions change.

Existing books provide valuable accounts of EU ETS history, market microstructure, and econometrics [4–6], while review-method guidance clarifies review design and reporting expectations [7,8]. Tang et al. [9] map quantitative ETS research across broad model classes and agent tasks, with derivatives as one subtopic; Bisiotis et al. [10] primarily synthesize time-series, econometric, and machine-learning forecasts. Neither provides an MSR-aware comparison linking institutional design to pricing-measure choice, option-surface validation, and implementable hedging evidence. This paper therefore asks three linked questions: how modelling evolved from Black–76-type specifications to institutionally grounded models; through which mechanisms EU ETS rules affect risk-neutral dynamics and derivative values; and how strongly the available evidence supports comparative robustness across regimes. Section 2 defines the common valuation concepts, Section 3 synthesizes three connected strands of evidence, Section 4 proposes a conditional model-selection principle and identifies research gaps, and Section 5 answers the questions from the present synthesis.

2. Theoretical Background

Black–76 prices a futures option by discounting the expected terminal payoff under lognormal risk-neutral futures dynamics with constant volatility [1]. Its practical virtues are substantial: the inputs are observable or conventionally inferred, the formula is transparent, and implied volatility provides a common coordinate for comparing strikes and maturities. The embedded assumptions nevertheless form an audit checklist rather than a default description of the EUA market. Continuous trading, a single diffusion, stable volatility, frictionless hedging, and a correctly specified terminal payoff can each fail without making the quoted Black–76 volatility useless. The benchmark should therefore be retained as a reference price and reporting convention while its economic and statistical assumptions are tested.

A second distinction concerns probability measures. Historical returns, realized volatility, GARCH estimates, and policy-shock decompositions describe dynamics under the physical measure (P). Derivative valuation instead requires a risk-neutral measure (Q) or another selected pricing measure under which discounted tradable prices satisfy the relevant no-arbitrage restriction. Jumps, liquidity, non-traded compliance risk, and policy uncertainty may make the market incomplete, so the physical-to-pricing-measure transformation and the associated volatility or jump premia are part of the model rather than technical afterthoughts [2]. Forecasting success under the physical measure is consequently not equivalent to option-pricing or hedging validation.

Institutional rules enter at both the boundary and the state-transition levels. A compliance penalty and the probability of aggregate shortage influence terminal allowance value, while banking, borrowing, cancellation, and certificate substitutability determine intertemporal and cross-asset links [3–6]. Under the MSR, the Total Number of Allowances in Circulation (TNAC) can trigger intake, release, or cancellation, making future supply state-dependent. To compare heterogeneous studies consistently, this review uses IMME as a coding lens with four linked dimensions: institutional regime, model specification, probability measure, and empirical evidence. IMME is not proposed as a standalone theory. It is an analytical device for asking, for every model family, when it fails, why it fails, and what evidence defines its domain of use.

3. Literature Review

3.1 Review Design and Coding

This critical structured review synthesizes heterogeneous theoretical and empirical studies [7]. Searches through 11 August 2026 covered Web of Science, Scopus, ScienceDirect, SpringerLink, Wiley Online Library, Taylor & Francis Online, SIAM, Cambridge Core, and journal publishers, supplemented by backward and forward citation chaining. Search terms combined EU ETS, EUA, and emission allowance with futures, options, pricing, hedging, risk-neutral, jump, stochastic volatility, structural, equilibrium, MSR, and TNAC. Eligible sources were English-language peer-reviewed articles, academic books, and review-method papers relevant to allowance dynamics, derivative valuation, or institutional mechanisms. The retained corpus comprises 53 DOI- or publisher-verified records: 47 support the model-level synthesis and six provide contextual support, as identified in Appendix A. Original database exports and screening decisions were not preserved, so PRISMA screening counts are not reported [8]. The Institution–Model–Measure–Evidence (IMME) coding records institutional regimes and contract boundaries, model specifications, probability measures and pricing transformations, and empirical data, calibration, validation, and hedging evidence. Appendix A provides the coding protocol and an example.

3.2 Early Benchmarks and Model Deficiencies

Across the early literature, Black–76 is better understood as a transparent benchmark than as a complete data-generating model for EUA derivatives. Equilibrium evidence identifies volatility that changes with time and the allowance price [11], while mixture-GARCH analysis documents conditional heteroskedasticity and heavy tails [12]. Markov-switching evidence further shows that EUA dynamics can change abruptly across market states [13]. Most importantly, the prohibition on carrying Phase I allowances into Phase II created an institutional discontinuity between contracts that would otherwise appear to be adjacent maturities; jump-diffusion and convenience-yield extensions in early derivative work were therefore responses to a real compliance boundary rather than merely statistical embellishments [14].

The IMME comparison clarifies both the contribution and the limit of this strand. Under the physical measure, early spot and futures studies document conditional heteroskedasticity, heavy tails, and state changes that are inconsistent with a constant-volatility Gaussian process or a single stationary regime [11–13]. Daskalakis et al. [14] additionally show that a closed-phase banking prohibition changes the terminal and cross-maturity structure relevant to derivatives. These findings diagnose ingredients that a pricing model may need, but they do not by themselves identify risk-neutral dynamics or establish option-pricing superiority. The appropriate conclusion is not that every jump, regime, or stochastic-volatility extension dominates Black–76. It is that the benchmark's domain must be stated and that features should be added only when an institutionally matched mechanism and a relevant validation test justify them.

3.3 Institutionally Grounded Structural and Equilibrium Models

The next literature strand redefines an allowance as a contingent claim produced by the compliance system. In incomplete-market work, information about aggregate positions can generate jumps and hedging depends on a selected martingale measure rather than a unique one [15]. Stochastic-control and equilibrium models derive prices from firms' production, abatement, trading, and terminal compliance decisions [16,17], while market-design analysis shows that allocation and compliance rules affect equilibrium and rents [18]. A tractable risk-neutral model links terminal value to the statutory penalty and the pricing-measure probability of non-compliance [19], and jump-diffusion formulations translate discrete information or policy shocks into nonlinear valuation equations [20]. These approaches make the institutional boundary explicit, but many were designed for non-bankable or early-phase markets.

Structural models deepen that link by endogenizing state dynamics. Firm optimization under information asymmetry and intertemporal flexibility produces option values that need not resemble those from an exogenous lognormal process [21]. Forward-backward stochastic differential equation and partial differential equation systems connect demand for the polluting good, cumulative emissions, allowance prices, and derivatives across compliance periods [22]. The discontinuous compliance payoff can create a singular terminal condition whose numerical treatment affects option values and implied volatility [23]. Market-

consistent modelling of substitutable EUA and Certified Emission Reduction certificates likewise shows that eligibility rules restrict joint dynamics in ways a generic correlated-assets model may miss [24]. Economic interpretation improves, but latent emissions states, singular boundaries, parameter identification, and numerical approximation become additional sources of model risk.

Empirical and comparative studies partly arbitrate this trade-off. EUA futures-option evidence indicates that jump risk can materially change prices when combined with time-series dynamics and a stated pricing transformation [25]. Time-varying convenience yields and risk premia complicate joint calibration across maturities [26], while long-memory and regime comparisons caution against transporting one volatility law across market states [27]. An equilibrium model treating allowances as nonlinear claims on aggregate emissions reproduces nonstandard spot, futures, volatility, and option behaviour, although estimation is demanding [28]. Most directly, Hitzemann and Uhrig-Homburg [29] compare a geometric-Brownian-motion benchmark with six ETS-specific reduced-form models using daily Phase II EUA futures and futures-option prices from January 2008 to June 2012. Monthly calibrations are tested on the following month's option prices; two-period models perform best overall in 2011–2012, particularly out of sample, while added volatility flexibility sometimes worsens held-out performance. This supports institutional alignment within the tested Phase II models, with no transaction-cost-aware hedging or MSR-era test.

3.4 Modern Pricing and Validation Evidence in the MSR Era

The MSR changes the institutional baseline by making supply a feedback process rather than a fixed cap. Phase IV design choices and reserve thresholds alter expected scarcity and can be anticipated before implementation [30]. Private banking and risk aversion interact with temporal issuance, intake, and cancellation, so a current surplus can imply different future supply paths under different expectations [31]. Multiperiod singular systems extend the theory to several compliance dates with banking, borrowing, and withdrawal [32]. Fang et al. [33] compare lognormal (LN), skewness–kurtosis-modified lognormal (SKM), and two-component mixture-lognormal (MLN-2) models using weekly EUA call-option samples from Phase III (2017–2018) and Phase IV (2023). MLN-2 performs best overall across the two samples, although SKM is strongest in-sample in the short Phase IV window. Together with [29], this comparison supports performance gains within specific samples and regimes rather than universal model superiority. Its evidence remains limited by sparse contracts, low-liquidity quotes, and the absence of dynamic-hedging tests. Theory-guided transfer learning improves pricing for two Phase IV call contracts, but that narrow test set also remains preliminary [34].

A related body of evidence shows why pricing-measure claims must be separated from historical forecasting. Around the 2006 compliance event, physical and risk-neutral densities changed differently after verified-emissions disclosure [35], and realized-emissions announcements affected returns, volume, and volatility even when the calendar event was partly anticipated in option prices [36]. A time-varying variance risk premium confirms that implied variance is not a direct forecast of physical variance [37]. Forecasting studies themselves conflict: implied volatility contains forward information in one sample [38], whereas GARCH specifications perform better in another illiquid sample [39]. Jump evidence is equally conditional. High-frequency analysis identifies small and large jumps and questions a continuous Brownian component [40]; Bayesian estimation supports stochastic volatility with jumps [41]; yet systematic outlier treatment removes many apparent jumps [42]. Differences in sampling frequency, quote quality, liquidity, horizon, cleaning, and loss function explain why these findings should not be pooled into one universal ranking.

Post-MSR work improves institutional relevance without closing the validation chain. Long-run market models suggest that reserve reform can raise prices and reduce cumulative emissions while making the effective cap endogenous [43], and a critical review documents the reserve's theoretical channels and contested effects [44]. Structural shock decomposition indicates that the composition of EUA price movements changed after the MSR, although the pandemic and energy crisis impede clean causal attribution [45]. A recent multistage pricing framework makes TNAC thresholds, reserve intake, release, cancellation, auctions, and continuation value explicit and reproduces recent price and futures-term-structure features [46]. Recent jump-test option pricing also reinforces event sensitivity [47]. Both advances remain early: neither demonstrates a complete option-surface calibration together with genuinely held-out prices and transaction-cost-aware dynamic hedging. The missing link is thus bounded but important—MSR states have not yet been connected to risk-neutral futures and option surfaces and tested under one reproducible cross-regime protocol.

4. Discussion and Synthesis

4.1 Central Synthesis

The synthesis rejects a monotonic complexity narrative. Black–76 fails as a structural account when volatility, tails, phase boundaries, or state-dependent supply are material, but it remains valuable as an interpretable quotation and control model. Jump, stochastic-volatility, mixture, structural, equilibrium, and machine-learning specifications can correct identified deficiencies, yet every correction creates new exposure to parameter instability, latent-state identification, numerical choices, or data scarcity. A model is therefore defensible only conditionally: its institutional boundary must match the contract and evaluation period, its pricing measure must be explicit, and its incremental mechanism must be supported by evidence relevant to the intended decision.

Viewed through IMME, the most persuasive literature connects all four dimensions rather than excelling at one. Institutional coherence without an identified market process is not a pricing model; a sophisticated process under the wrong compliance regime is economically misspecified; a physical-measure forecast without a pricing transformation does not validate an option value; and a low calibration error without held-out or hedging evidence does not establish robustness. The available direct comparisons identify benefits within their particular samples and model sets [29,33], but they do not establish a phase-invariant ranking. The review therefore proposes conditional adequacy as a model-selection principle.

4.2 Applicability Boundaries and Evidence Hierarchy

Model choice should follow the task. Black–76 can serve as a first benchmark for liquid, near-the-money quotations when a transparent implied-volatility language is needed. Historical GARCH, regime, or jump models are useful for forecasting and risk scenarios, but cannot price options without a justified measure transformation. Jump, stochastic-volatility, and mixture models are better candidates for smiles, tails, and event exposure when option liquidity identifies their added parameters. Structural and equilibrium models are most valuable when compliance, banking, certificate links, or MSR feedback are central, although their implementation requires sensitivity analysis for latent states and numerical boundaries. Data-driven models should be judged only under time-ordered tests, transparent training data, and strong conventional controls.

The evidence hierarchy proposed in this review is a cumulative four-level scheme derived from recurring validation gaps across the three synthesis strands. Level 1 requires institutional, no-arbitrage, and pricing-measure coherence. Level 2 adds identified and stable in-sample calibration with disclosed quote and liquidity filters. Level 3 requires genuinely time-ordered held-out option pricing by strike and maturity. Level 4 adds transaction-cost-aware dynamic hedging and transport across policy regimes, including explicit recalibration or retirement criteria. Greater mathematical complexity does not raise a model's level; each level adds a validation burden and does not substitute for lower-level coherence. In the reviewed literature, evidence is concentrated at Levels 1–2, with fewer studies providing Level 3 validation and very limited Level 4 evidence. Heterogeneous contracts, filters, samples, and losses accordingly preclude a pooled ranking by raw error.

4.3 Future Research

Three gaps emerge from the reviewed evidence: comparable out-of-sample option pricing, dynamic hedging with transaction costs, and validation across policy phases. Shared contract samples, quote filters, and time-ordered evaluation windows would make pricing comparisons more informative. Hedging tests could then examine whether pricing improvements translate into gains after trading costs, while cross-phase tests would assess whether calibrated models remain useful when compliance rules or MSR mechanisms change. These priorities would strengthen the evidence for conditional model selection and help distinguish ordinary parameter instability from institutional misspecification.

5. Conclusion

The present synthesis answers the first research question by identifying an evolution from a transparent Black–76 convention to empirical jump, volatility, and regime extensions, and then to risk-neutral, equilibrium, structural, multiperiod, MSR-explicit, and theory-guided data-driven models. The decisive differences among

these families are not mathematical sophistication alone. They lie in the terminal boundary, institutional state variables, probability-measure treatment, parameter identification, and standard of validation.

For the second question, EU ETS rules affect derivative values through several connected mechanisms: banking and borrowing determine intertemporal survival; penalties and aggregate shortage shape terminal value; disclosure and compliance dates alter the information process; certificate eligibility constrains joint dynamics; and the MSR makes future supply depend on circulating allowances and policy thresholds. These channels can change both the expected futures path and the risk-neutral distribution of option payoffs, so institutional design cannot be relegated to an external policy discussion.

For the third question, the limited direct comparisons provide relatively stronger support for institutionally aligned specifications within matched samples, model sets, and policy regimes; evidence for a broader ranking, especially in the MSR era, remains insufficient. Robustness requires coherent institutional boundaries and an explicit pricing measure, followed by held-out option pricing, transaction-cost-aware dynamic hedging, and validation across policy regimes. The resulting model-selection principle is conditional adequacy: use the least complex model that is regime-aligned, measure-explicit, empirically identified, and validated for the intended pricing or hedging task.

References

- [1] Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics*, 3(1–2), 167–179. [https://doi.org/10.1016/0304-405X\(76\)90024-6](https://doi.org/10.1016/0304-405X(76)90024-6)
- [2] Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3), 519–547. <https://doi.org/10.1111/j.1467-9965.2006.00281.x>
- [3] Rubin, J. D. (1996). A model of intertemporal emission trading, banking, and borrowing. *Journal of Environmental Economics and Management*, 31(3), 269–286. <https://doi.org/10.1006/jeem.1996.0044>
- [4] Ellerman, A. D., Convery, F. J., & de Perthuis, C. (2010). *Pricing carbon: The European Union Emissions Trading Scheme*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139042765>
- [5] Chevallier, J. (2012). *Econometric analysis of carbon markets: The European Union Emissions Trading Scheme and the Clean Development Mechanism*. Springer. <https://doi.org/10.1007/978-94-007-2412-9>
- [6] Ibikunle, G., & Gregoriou, A. (2018). *Carbon markets: Microstructure, pricing and policy*. Palgrave Macmillan. <https://doi.org/10.1007/978-3-319-72847-6>
- [7] Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- [8] Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- [9] Tang, L., Wang, H., Li, L., Yang, K., & Mi, Z. (2020). Quantitative models in emission trading system research: A literature review. *Renewable and Sustainable Energy Reviews*, 132, 110052. <https://doi.org/10.1016/j.rser.2020.110052>
- [10] Bisiotis, K., Christopoulos, D., & Tzougas, G. (2026). Forecasting carbon prices: A literature review. *Journal of Forecasting*, 45(2), 496–529. <https://doi.org/10.1002/for.70054>
- [11] Seifert, J., Uhrig-Homburg, M., & Wagner, M. (2008). Dynamic behavior of CO₂ spot prices. *Journal of Environmental Economics and Management*, 56(2), 180–194. <https://doi.org/10.1016/j.jeem.2008.03.003>
- [12] Paoletta, M. S., & Taschini, L. (2008). An econometric analysis of emission allowance prices. *Journal of Banking & Finance*, 32(10), 2022–2032. <https://doi.org/10.1016/j.jbankfin.2007.09.024>
- [13] Benz, E., & Trück, S. (2009). Modeling the price dynamics of CO₂ emission allowances. *Energy Economics*, 31(1), 4–15. <https://doi.org/10.1016/j.eneco.2008.07.003>

- [14] Daskalakis, G., Psychoyios, D., & Markellos, R. N. (2009). Modeling CO2 emission allowance prices and derivatives: Evidence from the European trading scheme. *Journal of Banking & Finance*, 33(7), 1230–1241. <https://doi.org/10.1016/j.jbankfin.2009.01.001>
- [15] Çetin, U., & Verschuere, M. (2009). Pricing and hedging in carbon emissions markets. *International Journal of Theoretical and Applied Finance*, 12(7), 949–967. <https://doi.org/10.1142/S0219024909005531>
- [16] Carmona, R., Fehr, M., & Hinz, J. (2009). Optimal stochastic control and carbon price formation. *SIAM Journal on Control and Optimization*, 48(4), 2168–2190. <https://doi.org/10.1137/080736910>
- [17] Kijima, M., Maeda, A., & Nishide, K. (2010). Equilibrium pricing of contingent claims in tradable permit markets. *Journal of Futures Markets*, 30(6), 559–589. <https://doi.org/10.1002/fut.20430>
- [18] Carmona, R., Fehr, M., Hinz, J., & Porchet, A. (2010). Market design for emission trading schemes. *SIAM Review*, 52(3), 403–452. <https://doi.org/10.1137/080722813>
- [19] Carmona, R., & Hinz, J. (2011). Risk-neutral models for emission allowance prices and option valuation. *Management Science*, 57(8), 1453–1468. <https://doi.org/10.1287/mnsc.1110.1358>
- [20] Borovkov, K., Decrouez, G., & Hinz, J. (2011). Jump-diffusion modeling in emission markets. *Stochastic Models*, 27(1), 50–76. <https://doi.org/10.1080/15326349.2011.542726>
- [21] Chesney, M., & Taschini, L. (2012). The endogenous price dynamics of emission allowances and an application to CO2 option pricing. *Applied Mathematical Finance*, 19(5), 447–475. <https://doi.org/10.1080/1350486X.2011.639948>
- [22] Howison, S., & Schwarz, D. (2012). Risk-neutral pricing of financial instruments in emission markets: A structural approach. *SIAM Journal on Financial Mathematics*, 3(1), 709–739. <https://doi.org/10.1137/100815219>
- [23] Carmona, R., Delarue, F., Espinosa, G.-E., & Touzi, N. (2013). Singular forward-backward stochastic differential equations and emissions derivatives. *The Annals of Applied Probability*, 23(3), 1086–1128. <https://doi.org/10.1214/12-AAP865>
- [24] Barrieu, P., & Fehr, M. (2014). Market-consistent modeling for cap-and-trade schemes and application to option pricing. *Operations Research*, 62(2), 234–249. <https://doi.org/10.1287/opre.2013.1242>
- [25] Yang, S. S., Huang, J.-W., & Chang, C.-C. (2016). Detecting and modelling the jump risk of CO2 emission allowances and their impact on the valuation of option on futures contracts. *Quantitative Finance*, 16(5), 749–762. <https://doi.org/10.1080/14697688.2015.1059953>
- [26] Trück, S., & Weron, R. (2016). Convenience yields and risk premiums in the EU-ETS—Evidence from the Kyoto commitment period. *Journal of Futures Markets*, 36(6), 587–611. <https://doi.org/10.1002/fut.21780>
- [27] Segnon, M., Lux, T., & Gupta, R. (2017). Modeling and forecasting the volatility of carbon dioxide emission allowance prices: A review and comparison of modern volatility models. *Renewable and Sustainable Energy Reviews*, 69, 692–704. <https://doi.org/10.1016/j.rser.2016.11.060>
- [28] Hitzemann, S., & Uhrig-Homburg, M. (2018). Equilibrium price dynamics of emission permits. *Journal of Financial and Quantitative Analysis*, 53(4), 1653–1678. <https://doi.org/10.1017/S0022109018000297>
- [29] Hitzemann, S., & Uhrig-Homburg, M. (2019). Empirical performance of reduced-form models for emission permit prices. *Review of Derivatives Research*, 22(3), 389–418. <https://doi.org/10.1007/s11147-018-09152-7>
- [30] Perino, G., & Willner, M. (2017). EU-ETS Phase IV: Allowance prices, design choices and the market stability reserve. *Climate Policy*, 17(7), 936–946. <https://doi.org/10.1080/14693062.2017.1360173>
- [31] Tietjen, O., Lessmann, K., & Pahle, M. (2021). Hedging and temporal permit issuances in cap-and-trade programs: The Market Stability Reserve under risk aversion. *Resource and Energy Economics*, 63, 101214. <https://doi.org/10.1016/j.reseneeco.2020.101214>

- [32] Chassagneux, J.-F., Chotai, H., & Crisan, D. (2023). Modelling multiperiod carbon markets using singular forward–backward SDEs. *Mathematics of Operations Research*, 48(1), 463–497. <https://doi.org/10.1287/moor.2022.1269>
- [33] Fang, M., Tan, K. S., & Wirjanto, T. S. (2024). Valuation of carbon emission allowance options under an open trading phase. *Energy Economics*, 131, 107351. <https://doi.org/10.1016/j.eneco.2024.107351>
- [34] Chen, L., Feng, L., Wu, W., Jin, C., & Zhang, Y. (2026). Deep carbon option: A transfer learning method for carbon derivatives valuation based on option pricing theory. *Journal of Cleaner Production*, 560, 148318. <https://doi.org/10.1016/j.jclepro.2026.148318>
- [35] Chevallier, J., Ielpo, F., & Mercier, L. (2009). Risk aversion and institutional information disclosure on the European carbon market: A case-study of the 2006 compliance event. *Energy Policy*, 37(1), 15–28. <https://doi.org/10.1016/j.enpol.2008.07.030>
- [36] Hitzemann, S., Uhrig-Homburg, M., & Ehrhart, K.-M. (2015). Emission permits and the announcement of realized emissions: Price impact, trading volume, and volatilities. *Energy Economics*, 51, 560–569. <https://doi.org/10.1016/j.eneco.2015.07.007>
- [37] Chevallier, J. (2013). Variance risk-premia in CO2 markets. *Economic Modelling*, 31, 598–605. <https://doi.org/10.1016/j.econmod.2012.12.017>
- [38] Viteva, S., Veld-Merkoulova, Y. V., & Campbell, K. (2014). The forecasting accuracy of implied volatility from ECX carbon options. *Energy Economics*, 45, 475–484. <https://doi.org/10.1016/j.eneco.2014.08.005>
- [39] Byun, S. J., & Cho, H. (2013). Forecasting carbon futures volatility using GARCH models with energy volatilities. *Energy Economics*, 40, 207–221. <https://doi.org/10.1016/j.eneco.2013.06.017>
- [40] Chevallier, J., & Sévi, B. (2014). On the stochastic properties of carbon futures prices. *Environmental and Resource Economics*, 58(1), 127–153. <https://doi.org/10.1007/s10640-013-9695-2>
- [41] Kim, J., Park, Y. J., & Ryu, D. (2017). Stochastic volatility of the futures prices of emission allowances: A Bayesian approach. *Physica A: Statistical Mechanics and Its Applications*, 465, 714–724. <https://doi.org/10.1016/j.physa.2016.08.036>
- [42] Dutta, A. (2018). Modeling and forecasting the volatility of carbon emission market: The role of outliers, time-varying jumps and oil price risk. *Journal of Cleaner Production*, 172, 2773–2781. <https://doi.org/10.1016/j.jclepro.2017.11.135>
- [43] Bruninx, K., Ovaere, M., & Delarue, E. (2020). The long-term impact of the market stability reserve on the EU emission trading system. *Energy Economics*, 89, 104746. <https://doi.org/10.1016/j.eneco.2020.104746>
- [44] Borghesi, S., Pahle, M., Perino, G., Quemin, S., & Willner, M. (2023). The Market Stability Reserve in the EU Emissions Trading System: A critical review. *Annual Review of Resource Economics*, 15, 131–152. <https://doi.org/10.1146/annurev-resource-111820-030145>
- [45] Maulberger, A., & Rathgeber, A. W. (2026). Disentangling supply and demand shocks in the EU ETS—Before and after the introduction of the Market Stability Reserve. *Resource and Energy Economics*, 87, 101575. <https://doi.org/10.1016/j.reseneeco.2026.101575>
- [46] Chen, T.-F., Chiu, H.-Y., & Shen, C. (2026). Modeling EUA price dynamics: A multistage PDE framework with Market Stability Reserve mechanisms. *International Review of Economics & Finance*, 110, 105626. <https://doi.org/10.1016/j.iref.2026.105626>
- [47] Liang, X., & Pan, Y. (2026). Research on pricing of carbon options based on an improved jump test method. *Finance Research Letters*, 90, 109406. <https://doi.org/10.1016/j.frl.2025.109406>
- [48] Chevallier, J. (2010). Modelling risk premia in CO2 allowances spot and futures prices. *Economic Modelling*, 27(3), 717–729. <https://doi.org/10.1016/j.econmod.2010.01.012>

- [49] Chevallier, J., & Sévi, B. (2011). On the realized volatility of the ECX CO2 emissions 2008 futures contract: Distribution, dynamics and forecasting. *Annals of Finance*, 7(1), 1–29. <https://doi.org/10.1007/s10436-009-0142-x>
- [50] Bredin, D., & Parsons, J. (2016). Why is spot carbon so cheap and future carbon so dear? The term structure of carbon prices. *The Energy Journal*, 37(3), 83–108. <https://doi.org/10.5547/01956574.37.3.dbre>
- [51] Hinz, J., & Novikov, A. (2010). On fair pricing of emission-related derivatives. *Bernoulli*, 16(4), 1240–1261. <https://doi.org/10.3150/09-BEJ242>
- [52] Kollenberg, S., & Taschini, L. (2016). Emissions trading systems with cap adjustments. *Journal of Environmental Economics and Management*, 80, 20–36. <https://doi.org/10.1016/j.jeem.2016.09.003>
- [53] Perino, G., & Willner, M. (2016). Procrastinating reform: The impact of the Market Stability Reserve on the EU ETS. *Journal of Environmental Economics and Management*, 80, 37–52. <https://doi.org/10.1016/j.jeem.2016.09.006>

Appendix A. Corpus Classification and IMME Coding Protocol

Corpus classification. The 53-record evidence matrix contains 47 records used in the model-level synthesis and six contextual records retained for triangulation. The latter comprise Chevallier's risk-premium analysis [48], Chevallier and Sévi's realized-volatility forecast [49], Bredin and Parsons' cost-of-carry and term-structure evidence [50], Hinz and Novikov's fair-pricing principles [51], Kollenberg and Taschini's cap-adjustment design [52], and Perino and Willner's analysis of delayed MSR reform [53]. They support pricing-measure, term-structure, or institutional interpretation but do not add an independent cross-model option-validation or hedging test; none was treated as a duplicate or quality exclusion.

Coding protocol. Institution codes record regime boundaries and policy state; Model codes identify processes, state variables, payoffs, parameters, and numerical methods; Measure codes distinguish P, Q, transformations, and incomplete-market pricing principles; Evidence codes record data, filters, validation splits, losses, hedging costs, and regime tests. For example, Fang et al. [33] is coded as Phase III–IV open trading with inter-phase banking (Institution); LN, SKM, and MLN-2 futures-option specifications (Model); risk-neutral calibration to option prices (Measure); and weekly cross-sections with training/testing MSE, MAE, and MAPE but no dynamic-hedging or regime-transport test (Evidence).

Illustrative evidence links. Risk-neutral modelling in [19] illustrates Level 1 coherence; option-price calibration in [33] illustrates a Level 2 component; and next-month pricing in [29] illustrates the chronological holdout component of Level 3. These links identify particular evidence requirements rather than formal ratings of whole studies; neither empirical example establishes the combined hedging and cross-regime validation required at Level 4.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Open Access

This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate

credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

