WORLDQUANT UNIVERSITY
Capstone

ORIGINAL RESEARCH
dated: May 20, 2024
doi: tbc

Forecasting carbon allowance prices across five emissions trading systems using deep learning (2017-2023)

https://github.com/pranavramkumar/wqu_capstone_5481/

Pranav Ramkumar[1], Artis J. Johnson[2] and Ekoue Jean Kougnah[3]

[1]*WorldQuant University*
[2]*WorldQuant University*
[3]*WorldQuant University*
[1]*pranavramkumar@gmail.com*, [2]*jammervp@gmail.com*, [3]*kougnahe@gmail.com*

ABSTRACT

The field of carbon pricing is a subtopic within environmental economics, which seeks to place a dollar price on the economic externality that is excess emissions by companies pursuing economic goals. Two main instruments of carbon pricing used by regulators have been carbon taxes and carbon permits (or allowances). Carbon allowances are permits to emit, which trade on an exchange (known as an emissions trading system or ETS) and can be bought or sold at a price, akin to a financial security. Researchers have been using statistical, machine learning or hybrid models to forecast the carbon price since the early 2000s. This paper contributes to the literature on carbon price forecasting in a few ways - using a feature set comprising human demographic, institutional, market, technological and external factors, forecasting allowance prices across 5 emissions trading systems for a comparative view, contrasting the performance of two of the most popular model architectures using in this domain namely CEEMDAN-LSTM and boosting models (specifically XGBoost), and finally introducing the popular transformer model, not widely used in the domain of carbon price forecasting to forecasting allowance prices across the same five ETS. The paper also presents a discussion around how the forecasts may be used to construct portfolios of carbon allowances to thereby securitize an exchange traded fund.

Results show that deep-learning based forecasting methods can provide a stable, effective and policy relevant means to forecast the carbon price. However, further investigation is required to improve model performance, and to address practical concerns to creating a commercial ETF backed by these price forecasts.

KEYWORDS: environmental economics, emissions control, carbon price forecasting, machine learning, deep learning, CEEMDAN, LSTM, Boosting, XGBoost, Transformer, ETF

JEL Codes: Q51, Q52, Q58, C6, Y80

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

LIST OF APPENDICES

1. INTRODUCTION

This section provides background on the history of carbon pricing and emissions control and a literature review of present-day emissions trading systems and prior efforts in research to predict carbon allowance prices since 2008.

1.1 Background

The Field of Carbon Pricing

Economics is the study of scarcity which involves two main types of problems - utility maximization subject to a budget constraint and cost minimization subject to output constraint. Environmental economics applies economic modeling to either explaining how a market system works (e.g. why is there pollution?) or address how an existing system should work or be optimally managed (e.g. what is the optimal level of pollution?). Carbon pricing is a sub-topic within environmental economics which seeks to place a price on the negative externality which is CO_2 emissions.

Researchers have raised arguments seeking to quantify biocentric views into economics since the 1830s. Robert Malthus, in his book ‘An Essay on the Principle of Population’, observed that humans had a propensity to use abundance for population growth rather than for maintaining a high standard of living - which he called the ‘Malthusian trap’; Populations had a tendency to grow until the lower class suffered hardships, want, war, famine and disease. He hence proposed that there must be limits to population growth, using an empirical model where population grew exponentially and food supply at best grew linearly. This set the tone to the biocentricity / anthropocentricity embedded in the field of environmental economics, and is the foundation for environmental concepts like ‘peak oil’, ‘peak minerals’ or ‘finite stock models’. Arthur Cecil Pigou, defined the concept of negative economic externalities (when an action taken by one economic agent directly enters the utility / profit function of another), and defined a Pigouvian tax, which increased a company’s private marginal cost of production upwards to the social marginal cost of production, which forced companies to produce at a social / environmental optimum rather than an economic / profit maximizing optimum, so as to eliminate externalities like pollution. This became the foundation of a carbon tax as a carbon pricing instrument. The figure below illustrates a ‘Pigouvian tax’ which is the distance between the private marginal cost of production and the social marginal cost of production lines in the chart below (equal to $3). Ronald Coase in his 1960 paper, ‘The Problem of Social Cost’ defined the ‘Coase Theorem’ which gave the methods for an efficient economic allocation in the presence of externalities, through the exploration of historical legal disputes between individuals and companies where the production function of one market agent influenced the utility of another.

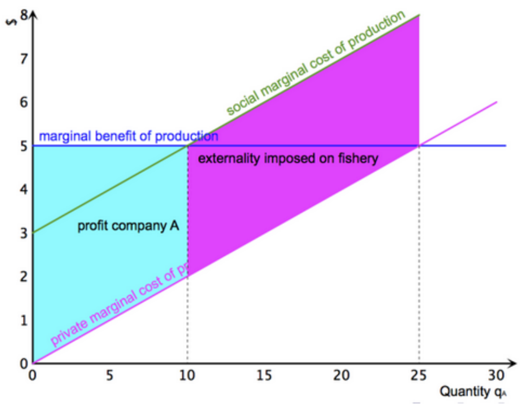


Figure 1 - Pigouvian Tax

Modern researchers have expanded upon the work of early researchers to bring specialized knowledge to the field of carbon pricing. Thomas H. Tietenberg, in 2010, wrote 'Cap and Trade: The evolution of an economic idea' and 'The Evolution of Emissions Trading', which transposed the academic theories of Coase and Pigou and outlined the mechanics of how an emissions control system could function. In the former paper, citing the works of other researchers namely John Dales (an economist who in 1968 proposed a permit system for managing water pollution in the Great Lakes), Tom Crocker (an economist who in 1966 wrote "In short, the job of the formulator of air pollution control policy is to bring about that correspondence between individual and collective welfare-however defined-which is lacking because the individual cannot or will not be operationally cognizant of the connection between certain of his means of achieving and maintaining his objectives and the achievement and maintenance of certain objectives for the collection of individuals" (Crocker 218)), Baumol and Oates (who in 1971 proposed a "charge system as opposed to a marketable permits system designed to meet a predetermined environmental target, with perfectly equivalent mathematics as with emissions trading" (Tietenberg 360)), and Montgomery (who in 1972 proved the existence of a cost-effective permit market equilibrium accounting for the location of emissions. "In general, those sources having higher marginal impacts on the environmental target need to pay higher prices per unit of emissions, which can be implemented by having separate permits for each receptor location" (Tietenberg 360)), he concluded environmental control efforts in the U.S., prior to emissions trading, centered around command-and-control policies since governors were not in favor of halting economic production for reasons of environmental wellness. In the latter paper, he proceeded to provide the full workings for an emissions trading system and concluded that it succeeded because it created incentives that were compatible with achieving a predetermined environmental target, at minimum cost, in the absence of any regulatory information on control costs.

Since the establishment of International and U.S. Pollution control accords starting with the Clean Air Act of 1963, there have been several different kinds of scholarship additionally brought to the field of carbon pricing and emissions control by economists and climate scientists. Michael Greenstone emerged as a key advisor to economic and climate policies in the U.S. in the past decade. His 2001 paper titled 'The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the census of manufactures', was a comprehensive record of how an air-quality regulation mechanism in the U.S. affected market participants. (The Clean Air Act, introduced in 1963 and amended in 1970, 1977 and 1990 provided the first U.S. federal regulatory system around carbon monoxide (CO), tropospheric ozone (O_3), sulphur dioxide(SO_2)), and total suspended particulates (TSPs). The 1970 amendment led to the formation of the U.S. Environmental Protection Agency (EPA) as a regulatory body. The 1977 amendments embedded more carbon pricing into national regional and sub-regional policy and the 1990 amendment established the first 'pollution permits' market where emitters were required to purchase a permit for each emitted unit with the EPA serving as national regulator and setting the cap of allowable emissions.) The comprehensive study brought together, for the first time, power plant emissions data and county-level submissions from 12 industry sectors covered under the regulation. It also provided insight into the difference between the counties which were able to attain their abatement targets and the economic benefits thereof (on employment, per capita incomes), versus those that were not using econometric fixed-effects regressions ("nonattainment counties have higher population densities, rates of urbanization, average education levels, per capita income, and per capita government revenues. Moreover a smaller fraction of their jobs are in the manufacturing sector, and they have lower poverty rates" (Greenstone 14)). William Nordhaus won the Nobel Prize in 2018 for his quantitative modeling of the effects of climate policy interventions on macroeconomics. His quantitative models (called DICE (Dynamic Integrated Model of Climate and Economy) and RICE (Regional Integrated Model of Climate and Economy), are used in scenario generation and modeling the global economic and social costs from not meeting the goals set out in climate accords (such as the Paris Agreement). In his 1975 Working Paper titled 'Can we control Carbon Dioxide?', he first drew links between the carbon cycle, how the atmospheric build up of CO_2 leads to heating and warming, and the impacts on social and economic outcomes (Nordhaus 5). His inspiration stemmed from other researchers on similar topics at the time, like Hubert Horace Lamb (H. H. Lamb) and Charles David Keeling (famous for the 'Keeling curve'). Scott Barrett (Lenfest-Earth Institute Professor of Natural Resource Economics at Columbia), is regarded as an industry expert in international cooperation around climate goals). He has participated and advised in the negotiation of several laws and accords pertaining to the governance of public goods. In a 1999 paper and book titled 'Montreal versus Kyoto', he compares the prevailing conditions in science and policy prior to the design and formulation of the Montreal Protocol (1987) (the treaty around the regulation of stratospheric ozone) and the Kyoto Protocol (1992) (the first treaty designed for carbon emissions control). The main differences highlighted are (i) 'measurability' - of stratospheric and atmospheric variables and the difference in maturity of the US EPA and the IPCC during the period, and (ii) 'unilateral vs. coordinated action' - while unilateral action by the U.S. to domestically abate CFC production was sufficient for success in the case of the Montreal protocol, a consensus of over 50% of all member states was required in the context of Kyoto protocol to be able to 'tip' the total level of carbon emissions to a sustainable trajectory (Barrett 197). Lawrence Goulder (Shuzo Nishihara Professor in Environmental and Resource Economics at Stanford University) has been an influential voice on environmental and economic impacts of environmental policies in the U.S. and China, with a focus on policies to deal with climate change and air pollution. He has been an influential voice in advising and studying the design of China's ETS pilots. Wolfram Schlenker, Co-director Center for Environmental Economics and Policy (CEEP), has been a leading environmental economist focused purely on emissions from agriculture, and from the reduction in crop yields from extreme weather, pollution and greenhouse gas accumulation. Other leading Macroeconomics such as Thomas Picketty, Paul Krugman, Joseph Stiglitz have also spoken about the role of carbon pricing in their policy proposals.

The large in-roads in the field of Artificial Intelligence and Machine Learning have come from the works of several researchers, who have brought the abilities to handle large datasets (exa and peta-byte scale) and large feature sets into the field of climate forecasting. Yoshua Bengio, a famous Canadian Computer Scientist, founder of the Montreal Institute of Learning Algorithms and recipient of the A. M. Turing Award (also called the Nobel Prize of computing, together with his collaborators Geoffrey Hinton and Yann Le Cun), is a leading specialist in Deep Learning architectures and Neural Networks and has been a leading voice in the development of many modern AI/ML literature in addition to several California based research groups like Facebook Research and Google DeepMind. He has recently served as an advisor (together with other pioneers in Deep Learning such as Andrew Ng (of Stanford) and Konrad Kording (of UPenn)) for Climate Change AI, a project which has grown to apply AI/ML and deep learning to several problems in the earth sciences, thereby leading to a flurry of research at the nexus of deep learning and climate science.

Emissions Control Mechanisms

The Kyoto Protocol was an international treaty adopted by United Nations member states in 1997 which acknowledged anthropogenic causes for global warming and committed to the reduction of greenhouse gas emissions through scientific consensus. It was the first treaty to propose market-based mechanisms for CO2 emissions reduction. The fist mechanism called the International Emissions Trading was a permit based regulation that allowed holders of permits to pollute legally. Clean Development Mechanism is a standards body for developing decarbonization projects in different industry sectors. Lastly, the Joint Implementation was where one country / party can develop abatement projects but transfer the environmental benefits recognized to other parties through legal provisions, allowing for abatement to proceed where it is best possible. However, it was not very successful because it was not widely adopted, lacked legal binding on states and had ambiguity in its mechanisms.

The Paris Agreement (adopted in 2015 at the UNFCCC COP 21 conference), the successor of the Kyoto Protocol, introduced its own mechanisms, entities, and standards for emissions trading and project development. This new protocol made climate commitments and emissions reduction targets legally-binding on states through the submission and tracking of nationally determined contributions (NDCs). The NDCs are a set of commitments self-identified by member states. These commitments require tangible quantitative period targets on how to lower six different greenhouse gases within their respective economies and on how to lower industrial footprints. Country, Regional or State Governments enforce these targets through two predominant instruments. The first being carbon taxes on certain sectors and the second being carbon permits, also called allowances or allotments, correspond to a level of permissible emissions by the purchasing party. These permits can be bought and sold and are traded on exchanges.

Today, there are 'carbon exchanges' or 'emissions trading systems' (ETS) where carbon allowances trade for a price quoted in $/tCO2e. As of March 31, 2024, there were 73 carbon pricing initiatives (tax/permit), 27 exchanges implemented and 5 exchanges under development. The collective of regulatory exchanges represent the 'Global Carbon Markets' and accounts for about 22% of all emissions regulated. An ETS originates each permit issued from the regulatory body managing that exchange, such as the EU Emissions Trading System, or the California Air Resources Board. A permit allows the purchaser or owner to pollute beyond the mandatory cap (in a cap and trade system) by the allowance owned. If an entity (corporate, government, fund etc..) does not have enough units of the allowance, it must purchase permits from another entity that has an excess of allowances. This characteristic, of being able to benefit from a permit or sell it, makes a permit akin to a fungible (traded) financial security.

INTERNATIONAL EMISSIONS TRADING (IET)

Article 17 of the Kyoto Protocol
Countries with commitments under the Kyoto Protocol can acquire emission units from other countries with commitments under the Protocol and use them towards meeting a part of their targets. An international transaction log, a software-based accounting system, ensures secure transfer of emission reduction units between countries.

The Kyoto Protocol spurred the creation of the European Union Emissions Trading Scheme, and many people foresee the growth and linking of emissions markets globally.

JOINT IMPLEMENTATION (JI)

Article 6 of the Kyoto Protocol
Through the JI mechanism, a country with an emission-reduction limitation commitment under the Kyoto Protocol may take part in an emission-reduction (or emission removal) project in any other country with a commitment under the Protocol, and count the resulting emission units towards meeting its Kyoto target.

JI projects earn emission reduction units (ERUs), each equivalent to one tonne of CO_2. As with the CDM, all emission reductions must be real, measurable, verifiable and additional to what would have occurred without the project.

Under JI there are two "tracks" by which projects can apply for approval: Party-varification and international independent body varification. The mechanism is overseen by the JI Supervisory Committee, which answers ultimately to the countries that have ratified the Protocol.

CLEAN DEVELOPMENT MECHANISM (CDM)

Article 12 of the Kyoto Protocol
The CDM allows emission-reduction (or emission removal) projects in developing countries to earn certified emission reduction (CER) credits, each equivalent to one tonne of CO_2. These CERs can be traded and sold, and used by industrialized countries to a meet a part of their emission reduction targets under the Kyoto Protocol.

The mechanism stimulates sustainable development and emission reductions, while giving industrialized countries some flexibility in how they meet their emission reduction limitation targets.

The projects must qualify through a rigorous and public registration and issuance process designed to ensure real, measurable and verifiable emission reductions that are additional to what would have occurred without the project. The mechanism is overseen by the CDM Executive Board, answerable ultimately to the countries that have ratified the Kyoto Protocol.

The mechanism is seen by many as a trailblazer. It is the first global, environmental investment and credit scheme of its kind, providing a standardized emissions offset instrument, CERs.

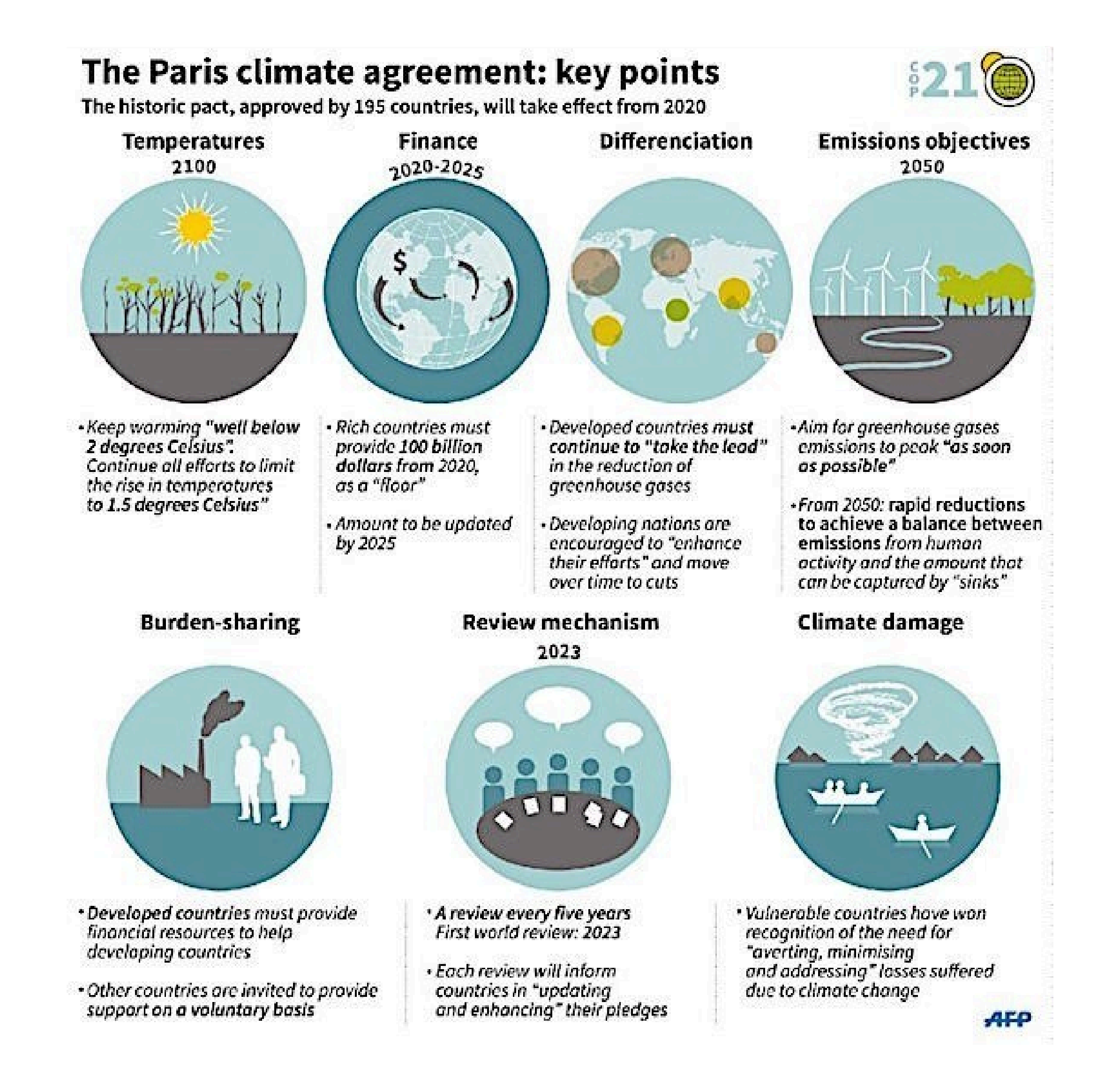
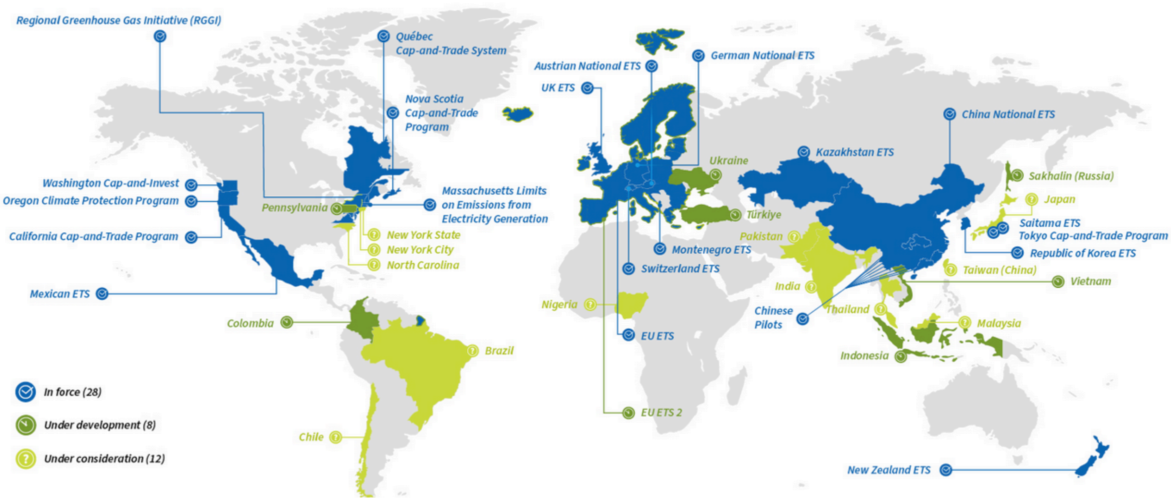


Table 1 - Kyoto Protocol and Paris Agreement Mechanisms
(Source: *UNFCCC and AFP***)**

1.2 Literature Review

Emissions Trading Systems

As described in section 1.1, 'emissions trading systems' (ETS) are regulatory exchanges where carbon allowances trade for a price quoted in $/tCO2e. As of March 31, 2024, there were 73 carbon pricing initiatives (tax/permit), 28 exchanges implemented, 11 exchanges under consideration, and 9 exchanges under development. This is summarized in the figure 2 below.



Under Consideration (11)	Argentina, Chile, EU ETS for buildings and road transport, Malaysia, New York State, Nigeria, Pakistan, Taiwan, Thailand, NYC, North Carolina,
In Force (28)	Austria, Canada (Nova Scotia), Canada (Quebec), China (Beijing pilot), China (Chonqing pilot), China (Fujian), China (Guangdong), China (Hubei), China (National ETS), China (Shanghai pilot), China (Tianjin), EU ETS, German National, Indonesia, Japan (Saitama), Japan (Tokyo), Kazakhstan, Korea ETS, Mexico ETS, Montenegro, New Zealand, Swiss ETS, United Kingdom, California Cap and Trade, Massachusetts, Oregon, RGGI, USA Washington
Under Development (9)	Brazil, Colombia, India, Japan, Russian Federation, Turkiye, Ukraine, Pennsylvania, Vietnam

Figure 2 - Emissions Trading Systems around the World
(Source: *ICAP Carbon Action***)**

This study intends to make general claims about the best modeling approaches for carbon allowance price forecasting, and hence specifically considers five global exchanges on which approximately 98% of all allowances are traded on a daily basis. The emissions trading systems considered are the U.S. Regional Greenhouse Gas Initiative (RGGI) founded by a consortium of 10 states in the U.S. East Coast, the California Cap-and-Trade exchange governed by the California Air Resources Board (CARB) and part of the Western Climate Initiative (WCI) coalition since 2007, the EU-ETS of the European Union, Korea's National ETS set up by the Korean Ministry of Environment, and China ETS (which in the context of this paper is approximated using only the Beijing ETS pilot and Shanghai ETS pilot systems due to their large traded volumes). This study will seek to predict the allowance prices for these five ETSs, using a unique feature set per ETS (described later); A detailed review of the history of these five ETSs is hence conducted below.

The Regional Greenhouse Gas Initiative
The Regional Greenhouse Gas Initiative (RGGI), was founded in 2009 by a consortium of 12 U.S. states on the east coast namely Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, Vermont and Virginia (Virginia and Pennsylvania have since withdrawn from the RGGI in 2023 due to insufficient CO_2 emissions control budget, and the inability to issue new allowances). Each of the above states delegated to RGGI Inc. (a 501c3 non-profit and governing body of the RGGI), the duty to administer CO_2 allowances, and monitor emissions compliance under each state's CO_2 Budget Trading Program for qualifying entities. RGGI is a market-based cap-and-invest initiative. While it is the goal of each state to achieve their independently defined climate targets, the RGGI provides a medium for acquiring or selling allowances.

Currently, only ‘fossil-fuel-fired electric power generators with a capacity of 25 MW or greater (or 15 MW or greater in New York State)’ are covered by the regulation. They are required to hold allowances issued by any participating state, purchased from a regional auction or through secondary markets, equal to their CO_2 emissions over a three-year control period. Each allowance allows the regulated entity to emit one short ton of CO2 and the cap for 2024 across all participating states is 157,184,044 CO_2 allowances. The cap is defined at the start of each control period or year based on the budget, and is adjusted based on a ‘model rule’ based on the banked allowances (following program reviews); It is progressively lowered by RGGI to bring regulated entities into compliance and into a practice of abating emissions as shown in the figure below.

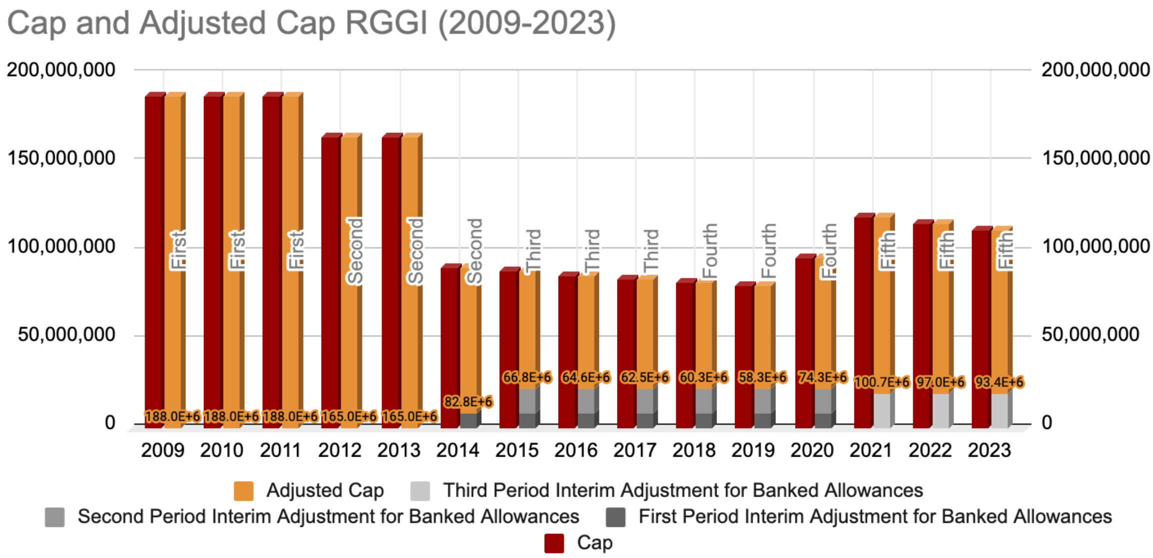


Figure 3 - RGGI Cap Setting in each Control Period
(Source: [RGGI])

As of the fifth control period (2021), 228 entities are regulated by RGGI, and the ETS covers ~14% of the emissions (631.8 $MtCO_2e$) of the participating states. Allowances are primarily sold in auctions at a price greater than or equal to the ‘minimum reserve price’ (equal to $2.56 in 2024). Of the 93 million allowances of 2023, ~92% were sold at an auction, a minimum amount at a fixed price, and of the remaining 8% (7.46 million), 5.5 million were sold from the cost containment reserve (a market stability mechanism of holding reserve allowances). Revenues from the sale of allowances are used by RGGI states for investing into energy efficiency, direct bill assistance, electrification, GHG abatement and clean and renewable energy projects. Entities are evaluated for compliance every three years (i.e. at the end of each control period).

The two market stability mechanisms of the RGGI are the CCR (Cost Containment Reserve) and the ECR (Emissions Containment Reserve) - the former is a reserve of allowances which are made available at auctions if the cost is likely to go beyond a threshold making the market unfair, the latter is a reserve of allowances which are withheld from auctions if the price falls below a threshold, in order to cap the level of total emissions, since purchasing the existing allowances is more accessible to market participants. In 2024, the CCR trigger price is $15.92 and the ECR trigger price is $7.35.

California Cap-and-Trade
California ETS (also known as California Cap-and-Trade) was launched in 2012 (compliance obligations began January 2013), as a mechanism for issuing, auctioning and trading CO_2 allowances for reducing GHG emissions for the state of California. California ETS program is the fourth largest in the world, after (EU-ETS, Korea and Guangdong) and its goal is to reduce GHG emissions of

covered industries 40% below the 1990 level before 2030. It is governed by the California Air Resources Board (CARB), and covers the power, industrial, transport, and buildings sectors.

Covered entities (~400 at present) must surrender allowances (from California ETS, or the Canadian Provinces of Quebec or Ontario where the program is linked to), for all their covered emissions. One allowance equals 1 $MTCO_2e$ and the cap for 2024 was 139,935,963 CO_2 allowances (estimated as total Vintage 2024 allowances). The cap is defined at the start of the compliance period and the number of allowances distributed is determined each year; Each year, fewer allowances are created and the annual cap declines. This brings regulated entities into compliance and into a practice of abating emissions as shown in the figure below.

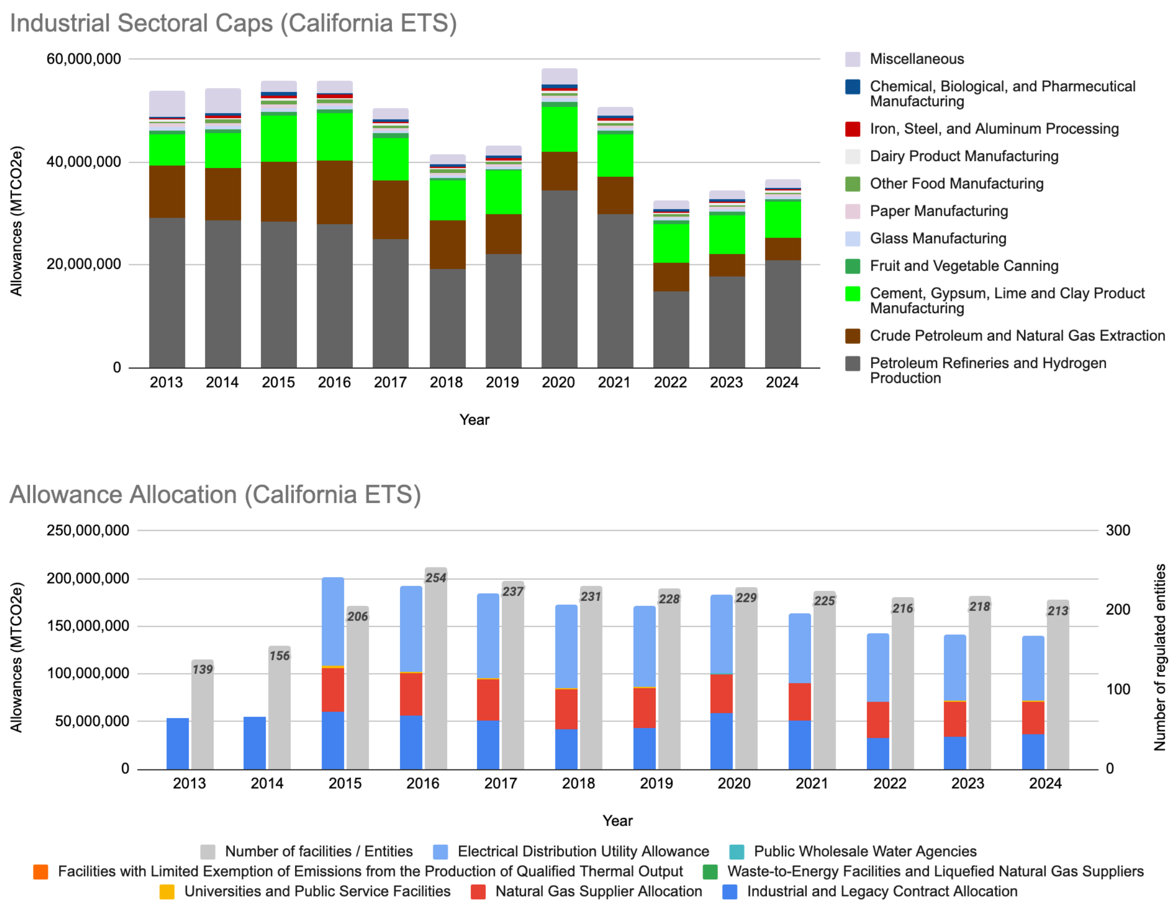


Figure 4 - California Allowance Allocation
(Source: *CARB*)

As of the fourth compliance period (2021), ~400 entities are regulated by California Cap-and-Trade and the ETS covers approximately 76% of the state's GHG emissions (292.20$MtCO_2e$). Allowances are distributed via free allocation (for industrial facilities based on production, cap adjustment and 'leakage risk'), free allocation with consignment (electricity distributors and natural gas suppliers), and auctions (where allowances are sold at a minimum price (USD 24 in 2024), or greater). Of all the Vintage 2023 allowances (281,955,193 in total), 70% were sold at an auction and the remaining were placed into reserve-tiers of market stability mechanisms. Revenues from allowance sales in auctions (~USD 27 billion to date) go into the GHG Reduction fund of which 35% must benefit disadvantaged and low-income communities. The funds are distributed via California Climate Investments into environmental, economic, and public health projects across the state. Entities are evaluated for compliance annually.

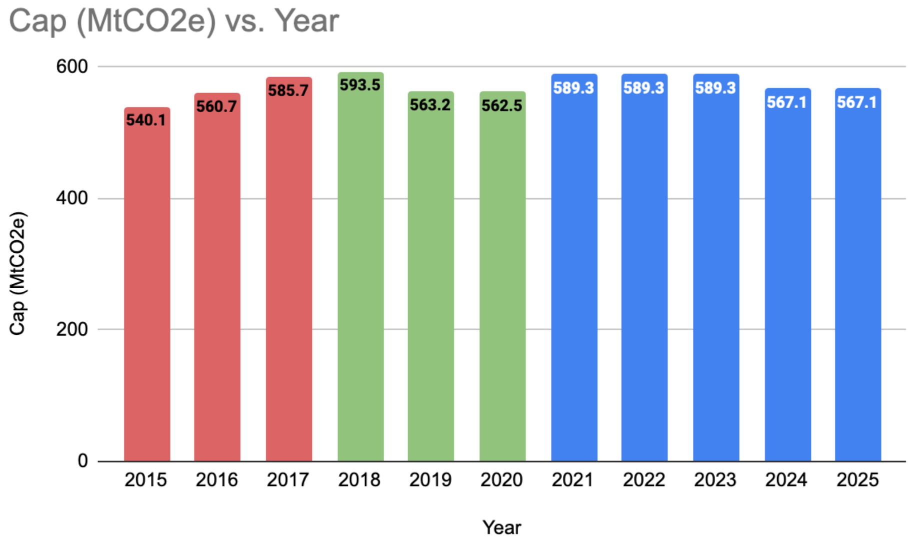
The main market stability mechanism of California Cap-and-Trade is the Allowance Price Containment Reserve (APCR). Prior to 2021, there were three APCR tiers, and after 2021 there are two APCR tiers and a price ceiling. The table below describes how allowances are assigned to these reserve tiers. Allowances from the reserve are made available for sale, if the price hits the threshold / ceiling, thereby making the market unfair.

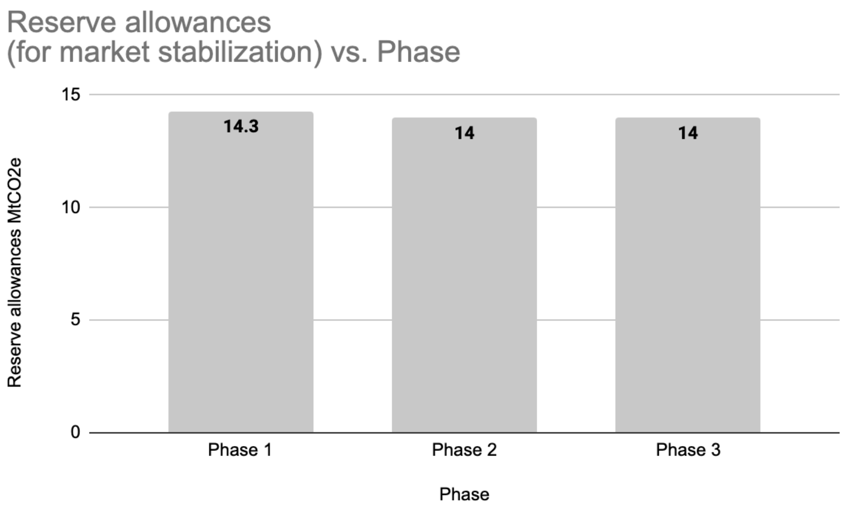
Prior to 2021			**After 2021**		
Tier 1	$\$x_1 - \$x_2$	Remaining 1/3rd of allowances from annual caps	**Tier 1**	$\$x_1 - \$x_2$ USD 56.20 for 2024	2/3rd of all allowances from annual caps + a portion of allowances from 2021-2030 annual caps.
Tier 2	$\$x_2 - \$x_3$		**Tier 2**	$\$x_2 - \$x_3$ USD 72.2 for 2024	
Tier 3	$\$x_3 - \$x_4$		**Price Ceiling**	$\$x_3 - \$x_4$ USD 88.2 for 2024	Unsold allowances from auctions

Table 2 - California ETS Market Stability Mechanism
(Source: *ICAP Carbon*)

Korea ETS
Korea ETS (K-ETS) was launched in January 2015 following the enactment of the 'Emission Trading Scheme Act' of 2012, and the launch of the First National GHG Inventory released by the Greenhouse Gas Inventory and Research Center (GIR) in 2012, and the First National GHG Inventory Management Plan in 2015. The goal of the ETS was to establish a market-based mechanism to meet the Paris Agreement 2030 NDC target (reducing GHG emissions by 40% from the 2018 level by 2030), and the country's objective to become carbon neutral by 2050. As East Asia's first nationwide, mandatory ETS, K-ETS is the second largest in the world. It is governed by the Ministry of Environment, which holds the overall responsibility for the K-ETS, the Ministry of Economy and Finance, which chairs the allocation committee, the Korea Exchange (KRX) which administers the trading auctioning platform and the Greenhouse Gas Inventory and Research Center (GIR), which is responsible for the registry, data and technical implementation.

Currently the Korea-ETS is in its third phase (2021-2025) and as of 2023, the K-ETS covers 804 entities from the maritime, waste, domestic aviation, transport, buildings, industry and power sectors. Covered entities can emit until a certain threshold, but must surrender KAUs (Korean Allowance Units) for every unit of emissions above the threshold (called 'covered emissions'). One KAU represents 1tCO_2e, and the cap for 2024 is of the order of 567.1 MtCO_2e. The cap is defined at the start of the compliance period and the number of allowances distributed is determined each year; Each year, technically, fewer allowances are created and the annual cap declines. However in reality, during the first phase of the exchange (2015-2017), the cap increased year on year, due to the issuance of new free allowances for new entrants, and also the sale of allowances from the market stability reserve for price control. The cap reduction and reserve mechanism is shown below.





Note: (i) Of the 1686.5 MtCO2e allowances in Ph 1, 89.4 MtCO2e was held in reserve for early action and new entrants. (ii) In Ph 2, 134 million for new entrants and 5 million for the market makers (iii) In Ph 3, 20 million is being held in reserve for for market makers

Figure 5 - K-ETS Emissions Cap and Reserve Allowances (Phase 1 - Phase 3) (Source: *ICAP Carbon***)**

As of 2021, K-ETS covered approximately 88.5% of all national emissions equal to 599MtCO_2e. Allowances are distributed via free allocation for select sectors such as grey clinker, oil refining, domestic aviation, waste, industrial parks, electricity generation, and district heating/cooling, and via auctioning for other sectors. At auctions, bidders can purchase up to 15% of auctioned allowances upwards of a minimum reserve price. However, as of 2023, only ~3% of the 589.3 MtCO2e 2023 cap, or 19 million allowances were auctioned. Revenues from allowance sales in auctions go into developing emissions mitigation infrastructure, low-carbon innovation, and technology development for small- and mid-sized companies covered by K-ETS.

The two types of market stability mechanism for K-ETS are triggers and instruments. Triggers represent very high or very low price levels where trading does not occur, and the allocation committee has to intervene with more free allowances from the reserve, or buy and retire some allowances. Instruments represent the use of other policy instruments or levers to alter allowance prices, such as additional ad-hoc auctions, limits on allowances held by a single entity, change in borrowing limit, offset limit, or the implementation of price ceilings or floors.

EU-ETS

The EU ETS was the first ETS launched in 2005, following the presentation of a paper ideating the design of the EU-ETS by the European Commission in Mar 2000, and the adoption of the EU ETS Directive in 2003, with the goal of reducing GHGs by ~55% from 1990 levels by 2030, and attaining climate neutrality by 2050. All EU-member states at the time became regulated by the EU-ETS beginning Phase 1 (2005-2007) of the ETS program. Three non-EU members (Iceland, Liechtenstein and Norway) joined the ETS in Phase 2 (2008-2012). Present membership includes the jurisdictions of Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden. While the EU-ETS is responsible for managing the regulatory framework and the administration of the EU registry, all member states have the obligation of implementing, verifying compliance with MRV and surrendering obligations.

As of present, entities regulated by the EU-ETS include energy and industry installations (8,640 as of 2022) and aviation and maritime transport companies (309 aircraft operators as of 2022). Each calendar year, per the Monitoring, Reporting and Verification regulations of the exchange, covered entities are expected to submit emissions monitoring plans, reports and have them verified by independent verifiers by the end of the year (March), and are expected to surrender one allowance per tCO_2e of covered emissions. The total emissions cap for 2024 is 1,442,572,462 tCO_2e. Over the past two phases 2013-2020 and 2021-2030, the cap has been reduced linearly by 1.74% and 2.2% each year. The cap reduction is shown in the figure below -

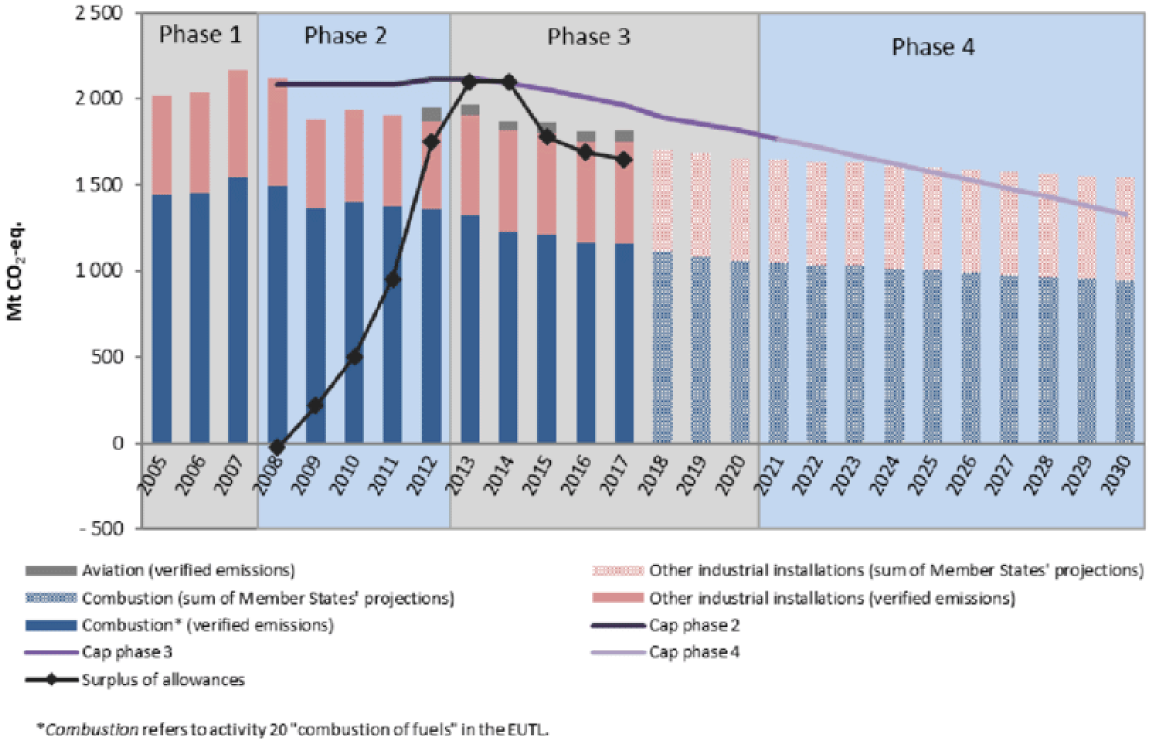


Figure 6 - EU-ETS Cap Reduction (Phase 1 - Phase 4)
(Source: *Amanatidis (2020)***)**

As of the start of Phase 4 (2021), despite its long history, the EU-ETS only covers ~38% of all EU emissions (1,335 $MtCO_2e$). This is due to the limited sectoral coverage and the nascency of emissions control in the Aviation sector (and whether it would be covered by the EU-ETS or other treaties such as CORSIA). As of 2023, a new ETS has been proposed for the buildings, road transport and additional sectors which is likely to increase emissions coverage. Similar to other ETS, allowances are distributed via free allocation for hard to abate sectors to cover 'leakage emissions' (emissions which could undo progress in other sectors), and every sector is assessed for technological progress towards decarbonization goals, to determine free allocation benchmarks. Other sectors participate in auctions (where allowances are sold at a price greater than or equal to a minimum price); As of Phase 4, the auctioning share of allowances was 57% of the cap, and hence the main method for issuing allowances in the context of EU-ETS. Revenues from allowance sales in auctions goes into the budgets of member states, 50% of which must be mandatorily used as state aid to climate and energy initiatives and reported to the European Commission annually.

Some market stability mechanisms in the context of EU-ETS (intended for price or emissions control) include (i) Backloading or withholding allowances from issue or sale for a period of 4-5 years by placing into the 'market stability reserves (MSR)' to reduce the allowances in circulation (ii) The 'Market Stability Reserve (MSR)' which either takes allowances away from circulation (24% if allowances > 1,096 million, or a lesser percentage if allowances > 833 million and < 1,096 million), or adds allowances to circulation (100 million added if allowances < 400 million) in order to influence price stability or emissions control.

China

Between 2013-2016, China launched 8 regional pilot ETS schemes with the view to discover - (i) If an ETS offers the right incentive for companies to reduce emissions, given the ambitious industrialization targets of China, and the impact of a Chinese slowdown on the global economy, (ii) The emissions cap and the carbon price that each region can sustain through a market based mechanism, and (iii) Which sectors can be regulated by an ETS and how allowances should be distributed and surrendered. The list of pilots are shown in the table below, rank ordered by the size of their emissions (and hence cap).

#	ETS	Founded	Sectoral Coverage	Cap	Avg. Auction Price in 2023 (USD)
	National	2021	Power	~5,000 $MtCO_2e$ (2022)	9.65
1	Guangdong	2013	Domestic Aviation, Industry	297 $MtCO_2e$ (2022)	10.58
2	Hubei	2014	Industry	180 $MtCO_2e$ (2022)	6.03
3	Fujian	2016	Domestic Aviation, Industry	116.2 $MtCO_2e$ (2022)	3.28
4	Shanghai	2013	Maritime, Domestic Aviation, Transport, Buildings, Industry, Power	100 $MtCO_2e$ (2022)	10
5	Chongqing	2014	Industry	78.39 $MtCO_2e$ (2022)	4.09
6	Tianjin	2013	Industry	74 $MtCO_2e$ (2023)	4.54
7	Beijing	2013	Transport, Buildings, Industry, Power	~44 $MtCO_2e$ (2022)	16.26
8	Shenzhen	2013	Transport, Buildings, Industry	28 MtCO2 (2023)	6.55

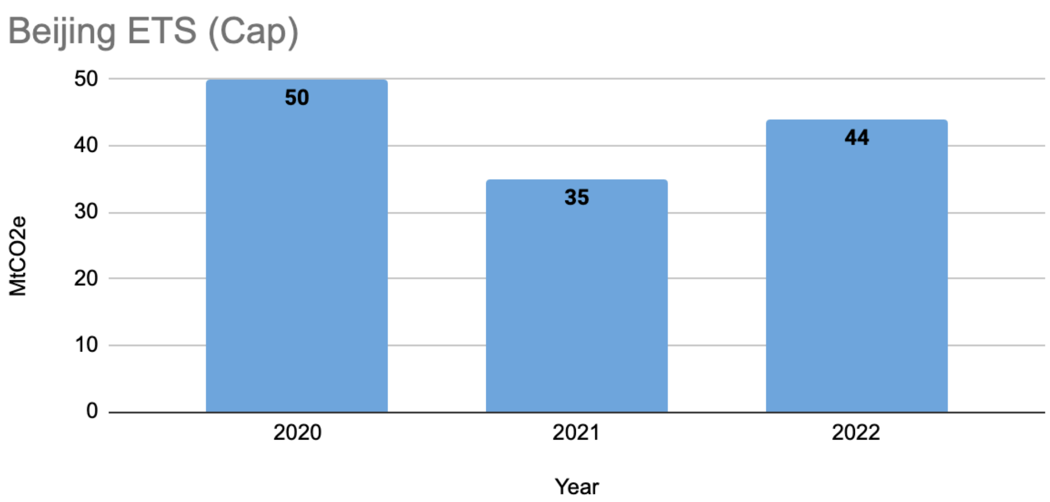
Table 3 - China's ETS Pilots (2013-2016)
(Source: *ICAP Carbon*)

As of 2021, a new National ETS was created to regulate the power sector. Due to the sheer scale of China, it became the world's largest ETS in terms of covered emissions (5 billion tCO_2e, 40% of National emissions). In order to cap the scale of the National program, as of 2024, the regional pilot schemes continue to remain in operation; For the purpose of this study, in order to have a comparable with the other ETSs with regard to average traded volumes and cross-sectoral coverage, we have chosen to work with only two pilot schemes namely - Beijing and Shanghai, which together with RGGI, California, Korea and EU-ETS account for ~98% of all allowances traded globally.

Cui et al (2021) performed an analysis of China's ETS pilot programs using firm emissions data obtained from the CNTSD, a large-scale annual survey conducted by China's Ministry of Finance and State Administration of Taxation. This database provides detailed firm-level information about energy consumption by source, including coal, natural gas, and electricity and economic information for a wide range of firms. The study intended to find if mass-based emissions control (based on total emissions), or rate-based emissions control (based on emissions intensity) is more effective, and if firms on average reduced their level of total emissions in absolute terms by complying with pilot ETS regulations. Using panel data from the period 2009-2015 and by means of matched difference-in-differences (DID) econometric approach, they concluded that the regional ETS pilots are effective in reducing firm emissions leading to a 16.7% reduction in total emissions and a 9.7% reduction in emission intensity.

Beijing Pilot ETS
Beijing Pilot ETS was launched in November 2013, to serve two functions - first, to serve as a regional ETS pilot for the Beijing metropolitan area and second, to play a supporting role in the establishment of the CCER (China Certified Emissions Reduction) national registry for carbon offsets. It is a relatively small ETS since it covers only ~30% of the city's emissions but is important by virtue of its trading volume and the level to which it is analyzed by country carbon experts. It is governed by the Beijing Municipal Ecology and Environment Bureau (EEB) (北京市生态环境局) and covers the Transport, Buildings, Industry and Power sectors. It has set itself the goal of reducing CO_2 emissions by 10% or more compared to 2030 modeled peak emissions, and by ~18% compared to 2020 measured emissions levels.

As of 2022, there were 909 entities which were covered and had to surrender allowances, and 388 entities which had mandatory reporting but did not need to surrender obligations. One allowance equals 1 tCO_2e and the cap as of 2022, was 44 MtCO2e. The cap is defined at the start of the calendar year, and entities are expected to surrender allowances by mid-June of the following year; Cap reduction of the Beijing ETS is shown in the figure below -



Notes: (i) 2021 is lower due to transfer of power sector to the national ETS
(ii) 2022 is higher due to 41 new covered entities in Beijing ETS (iii) Subsequent year data unavailable at the moment

Figure 7 - Beijing ETS Cap Reduction
(Source: *ICAP Carbon*)

Allowances are distributed via free allocation through grandparenting (based on historical emissions or emissions intensity in the past three years), or through benchmarking (specific allowance capacities for new entrants and sectors). Allowances (about 5%) are also auctioned (at a price greater than a minimum ~USD 16.26) through regular or ad-hoc auctions. Revenues from allowance sales in auctions (~USD 38.67 million to date, and USD 22.72 in 2023) go into the city treasury from where it is used for municipal projects.

The main market stability mechanisms of the Beijing ETS are - (i) A reserve mechanism managed by the authority where the EEB buys or sells allowances to maintain prices between CNY 20 (USD 3) and CNY 150 (USD 21) and (ii) Limits on price volatility (+-20% around the previous day average) beyond which trading for the day may be suspended and position size (annual allocated allowances + 50,000 - 1 million tCO_2e)

Shanghai Pilot ETS

[Shanghai Pilot ETS](#) was founded in November 2013 as China's second regional ETS (after Shenzhen), first, to serve as regional ETS for the Shanghai metropolitan area, and second, to serve as a hub for carbon finance innovation in China. It is a small ETS since it covers only ~36% of the city's emissions but is important by virtue of its trading volume and its ability to drive a perfect 100% compliance rate since its launch. It is governed by the Shanghai Municipal Ecology and Environment Bureau (EEB) (上海市生態環境局) and covers the industrial, buildings, aviation and maritime (as of 2016) sectors. It has set itself the goal of reducing CO_2 emissions by 5% compared to peak (2025) levels by 2035 and attaining full climate neutrality by 2060.

As of 2022, the Shanghai ETS had 357 covered entities and has recently added some data centers and logistics companies as eligible entities, the former with full compliance obligations and the latter with only reporting obligations. Entities have to surrender one allowance per unit of emissions, One allowance equals 1 tCO_2e and the cap as of 2022, was 100 $MtCO_2e$. Similar to the Beijing pilot, The cap is defined at the start of the calendar year, and entities are expected to surrender allowances by mid-June of the following year; Cap reduction of the Shanghai ETS is shown in the figure below -

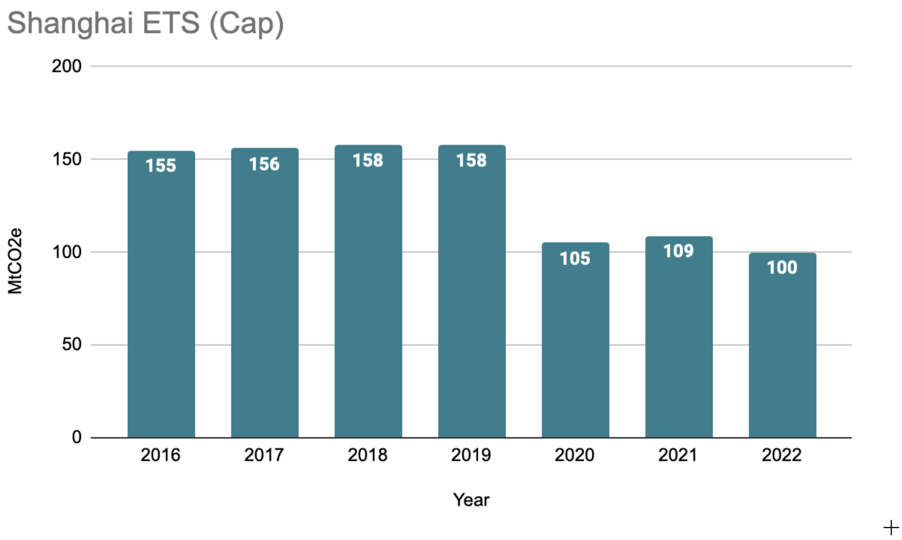


Figure 8 - Shanghai ETS Cap Reduction
(Source: *ICAP Carbon***)**

Allowances are distributed via free allocation through grandparenting (based on historical emissions or emissions intensity in the past three years for industrial sectors, aviation, ports, shipping and water suppliers), or through benchmarking (specific allowance capacities for new sectors namely electricity and heat generation, electricity grid, data centers). Allowances are also auctioned (at a price greater than a minimum ~USD 8-10) through two auctions per year since 2020. 3 million allowances sold for approx USD 10.5 per allowance in the November 2023 auctions. Revenues from allowance sales in auctions (~USD 27.03 million in 2023) go into the provincial treasury from where it is used for municipal projects.

In the case of Shanghai ETS, market stability mechanisms include (i) A share of annual allowances maintained in reserve for price control and (ii) Price stabilization through suspending trading or position size limits if price volatility exceeds 30% of prior day average.

For a tabular summary of the history and features of each of the above ETS, refer *Appendix A*.

Carbon Price Forecasting

Prior attempts by researchers to forecast carbon allowance prices can be roughly classified into three approaches - a statistical approach, a feature-focussed machine learning approach, and a hybrid modeling approach (Refer to *Appendix B* for a systematic review of the most popular papers since 2008).

The first approach referred to in this paper as the '*statistical approach*' (which has been used since as early as the first carbon exchange in 2005), involves working purely with the carbon allotment price signal, and applying time-series modeling techniques like ARIMA or GARCH variants. Benz and Truck (2008) used Markov transition, and AR-GARCH based modeling to forecast the log-returns of the EU-ETS allowance price wherein they concluded that it was important to assume a time and price dependent volatility for accurate predictions using GARCH, and that traders who viewed carbon allowances like commodity futures, were as interested in forecasting short-term intraday prices, as they were long-term spot prices. Chevallier (2008) used ARMAX-ARCH models to forecast returns of ECX carbon futures for Phase II (2008-2012) for EU-ETS using a set of exogenous covariates. He found that these statistical approaches could be beneficial because carbon futures returns exhibit non-zero skewness and excess kurtosis (fat tailed leptokurtic distribution). Byun and Cho (2013) used a suite of GARCH models (GARCH, EGARCH, TGARCH and GJR-GARCH) with normal and t-distributions and with AR, MA and ARCH terms varying from one to three to forecast EU-ETS EUA Carbon Futures Prices using a feature set. They concluded that based on loss functions such as MSE, MSE-LOG, MAE, MAE-LOG and QLIKE that the more complex GARCH specifications do not necessarily outperform simpler GARCH variations, and that residuals from the prediction under each case approach a normal distribution. Zhu and Wei (2013) showed that enhancing ARIMA models with Least Squares Support Vector Machines or Artificial Neural Networks could help predict the non-linear components of the carbon price adequately. Yang, et al (2021) in their 2021 paper which used the ICEEMDAN technique explained later in this paper, that "while statistical models were preferred by early researchers as they were available at the time, they were mainly linear in their analysis and results." This describes the limitation of the *statistical approach*, which led to the expansion to other approaches.

The second approach referred to in this paper as the '*feature-focused machine learning approach*' stresses the importance of identifying the right causal drivers of the carbon allowance price which generate the highest feature importances. Researchers have approached the problem of feature selection from different starting and vantage points, using informed priors or feature scoring techniques, and on average their opinions can be summarized in the following sentence - carbon prices are driven by *human factors, institutional factors, market factors, technological factors or external factors*. The *human factors* approach correlates carbon allowance prices with socio-economic phenomena such as social vulnerability and energy poverty. Shahzad et al (2023) used energy prices to proxy energy poverty in households and proposed features such as WTI NYMEX Futures, Natural Gas Futures and (Dow Jones) Heating Oil Futures, while Byun and Cho (2013) who made similar claims believed that energy volatilities were a better explanatory factor. Chevallier (2008) believed that the same phenomenon could be better described by observing dividend yields and the junk bond premium. Other approaches have suggested observing GDP/capita, inflation or markers of recession at the local level (state or sub-regional level) as an indicator of socio-economic stress. The *institutional factors* approach believes that the carbon allowance price is best predicted using covariates that identify how regulation and policy of the carbon exchanges shall be driven. This means studying the political affiliations (right or left wing) of ETS member states or countries, the presence of other policy instruments such as a carbon tax, transacted volumes, auctioned versus freely allocated allowances, fees and penalties, and the addition of new industry sectors into coverage. Researchers who believe that *market factors* have high explanatory power on the carbon allowance prices tend to believe that carbon markets are effectively commodities markets and are hence affected by similar macroeconomic factors such as equity market excess returns (Chevallier, 2008), industrial production and clean energy indices (Rudnik et al, 2022), commodity prices, financial market indices, macroeconomic variables like production, consumption and government spending, economic sentiment, and economic development indices (Wang et al, 2023). Peng et al (2023) focussing on predicting Shenzhen allowance prices (SZA) during the period Mar 2018 - Jul 2022, took inspiration from a work by Tan et al (2022) in the context of EU-ETS price prediction, and used an entire cohort of 53 explanatory variables including commodity variables (such as energy, non-ferrous metals, agri products), stock and bond market variables and other industry composite variables, upon which they performed dimensionality reduction using coefficient shrinkage and factor modeling techniques. Most researchers tend to concur that *technological factors,* such as the innate abilities of different industry sectors to decarbonize their operations are a good indicator of supply and demand for carbon allowances, and hence a good predictor of the price. Hence, tracking phenomena such as emission reduction goal setting, industry coverage additions to ETS programs, prevention, reduction and transition strategies of companies, and trends and activity in the voluntary carbon markets could be potentially good leading indicators of technological change. Further, there is unanimous consensus that *external factors* such as extreme weather (heat or cold), major climate events in the region or sub-region (Wang et al, 2023), and modern conflicts (Zhang et al, 2023) have explanatory power over carbon allowance prices.

The third forecasting approach referred to in this paper as the '*hybrid modeling approach*' involves a two-stage approach of signal-decomposition followed by prediction. The first stage of signal decomposition involves working purely with the 'carbon allowance price signal', decomposing it into a series of components which describe its trend and periodicity incrementally, reassembling it identifying elements of trend, jump, or pattern effects, and thereafter predicting the carbon allowance price. The most popular signal decomposition techniques employed in the first-stage in this approach are EMD (empirical mode decomposition), EEMD (ensemble empirical mode decomposition), MEMD (multi-variate empirical mode decomposition), VMD (variational mode decomposition), Fast-EMD (fast empirical mode decomposition), WT (wavelet transformation), SSA (singular spectrum analysis), CEEMDAN (Complete ensemble empirical mode decomposition with adaptive noise) or ICEEMDAN (Improved complete ensemble empirical mode decomposition with adaptive noise). The second stage involves predicting the reassembled carbon price vector or the decomposed signal components and thereby the carbon allowance price, using machine learning or deep learning models. Yang et al (2021) focusing on EU-ETS, Beijing ETS and Shenzhen ETS allowance price prediction used ICEEMDAN data pre-processing in combination with GBiLSTM, CNN and ELM (extreme learning machine) models and the multiobjective dragonfly optimization algorithm (MODA) for weighting sub-models into an ensemble. They obtained robust predictions with an MAE of 0.0005 for EU-ETS, 0.013 for Beijing ETS and 0.0448 for Shenzhen ETS. Lu and Ma (2020) found that CEEMDAN had the best predictive performance in predicting the 8 China pilot ETS allowance prices in combination with Radial Basis Function Neural Nets (RBFNN) or Grey Wolf Optimizer Arps decline model (GWO-KNEA) as second stage predictor. Yun et al (2022) concluded that CEEMDAN was so useful to the problem of carbon price forecasting due its 'multi-frequency' nature to mine for signal patterns at different time-scales (i.e. today's policy may not stem from yesterday's stress, but from a similar pattern of events 5 or 10 years ago). Wang, Jiang and Shu (2022) obtained an RMSE of

3.9 and an MAE of 0.0456 while forecasting China National ETS allowance prices, using singular spectral analysis in combination with LSTM.

In summary, the three approaches described above (and specifically the paradigms of mode decomposition, stochastic modeling, machine learning and deep learning), allow for capturing the characteristics innate to carbon allowance price forecasting, such as multi-frequency, latency, non-stationarity, non-linearity, variable memory, cross-domain feature sets, ensembling and cross-validation. For best results, most studies are done at the level of a single ETS, because explanatory factors may vary and sub-regional effects may dominate. Handling multiple ETS predictions simultaneously also is a difficult undertaking for reasons of full verifiability. While different hybrid model combinations show good performance, there is little room to compare different models outside the context of their studies, and the external validity for a model which has succeeded in one setting into another may not always be feasible.

This paper will specifically contribute to this literature on carbon price forecasting in a few ways - using a feature set which combines different factor approaches described earlier, making multi-market forecasts across 5 emissions trading system allowances, contrasting the performance of a transformer model and a boosting model (XGBoost) against the popular CEEMDAN-LSTM (used as benchmark) for carbon price forecasting and finally, reporting multiple loss functions and risk measures.

2. DATA

This section describes the different data sources pertinent to the carbon markets that are available, how the authors chose relevant features for the prediction problem at hand, and how the data was actually proprocessed for this study.

2.1 Data Availability

Several publicly available sources of carbon allotment prices across all the exchanges of the world are currently available for research. The World Bank Carbon Pricing Dashboard, provides data and insights on global carbon pricing initiatives, facilitating informed climate policy decisions (The World Bank also provides a famous communication guide for communicating the carbon price from emissions trading and carbon taxes in the context of other policy reforms). The ICAP Allowance Price Explorer provides a user-interface for comparing carbon pricing mechanisms, visualizing allotment prices and comparing different emissions trading systems. The Intercontinental Exchange (ICE), provides data on carbon allowance prices for various emissions trading schemes, including the EU Emissions Trading System (EU ETS) and the Regional Greenhouse Gas Initiative (RGGI) in the United States. European Energy Exchange (EEX), offers data on carbon allowance prices for the EU ETS and other European carbon markets. Carbon Market Data (CMD), is a platform that provides data and analytics on carbon markets worldwide, including carbon allowance prices, market trends, and regulatory developments. With regard to Voluntary Carbon Offset Projects, a database developed by the Berkeley Carbon Trading Project called the ‘U.C. Berkeley Voluntary Registry Offsets Database’, contains all carbon offset projects, credit issuances, and credit retirements listed globally by four major voluntary offset project registries—American Carbon Registry (ACR), Climate Action Reserve (CAR), Gold Standard, and Verra (VCS). These four registries generate almost all of the world's voluntary market offsets and also include projects eligible for use under the California / Quebec linked cap-and-trade programs. World Carbon Pricing Database, from Resources for the Future, explores carbon prices arising from carbon pricing mechanisms around the world since 1990. The tool allows users to explore the price associated with CO_2 emissions in national and subnational jurisdictions around the world between 1990 and 2021. It looks specifically at the two common types of carbon pricing mechanisms: carbon taxes, which set a price on carbon, and cap-and-trade systems, which caps the overall volume of emissions and creates a market for emissions allowances.

However, for commercial use-cases, the timing and veracity of the carbon price data available from online sources has been heavily challenged. Bloomberg Terminal, provides comprehensive financial data, including carbon allowance prices for various emissions trading schemes globally and live-data from carbon market auctions. Thomson Reuters Eikon, offers financial data and analytics, including carbon allowance prices for different carbon markets. ICIS provides market intelligence and pricing information for various commodities, including carbon allowances in different emissions trading schemes. Point Carbon (S&P Global Platts) offers market intelligence, analysis, and pricing data for carbon markets globally. Quantum Commodity Intelligence is a price reporting agency powering the global energy, biofuels, carbon and ammonia markets with price assessments, market news and trade data. Allied Offsets claims to be the world's largest aggregated data source for carbon offsetting. Carbon Dioxide Removals Data. Its CDR database has information over 845 carbon project developer companies around the world and their projects. Carbon Pulse, provides news and intelligence on carbon markets, greenhouse gas pricing, and climate policy.

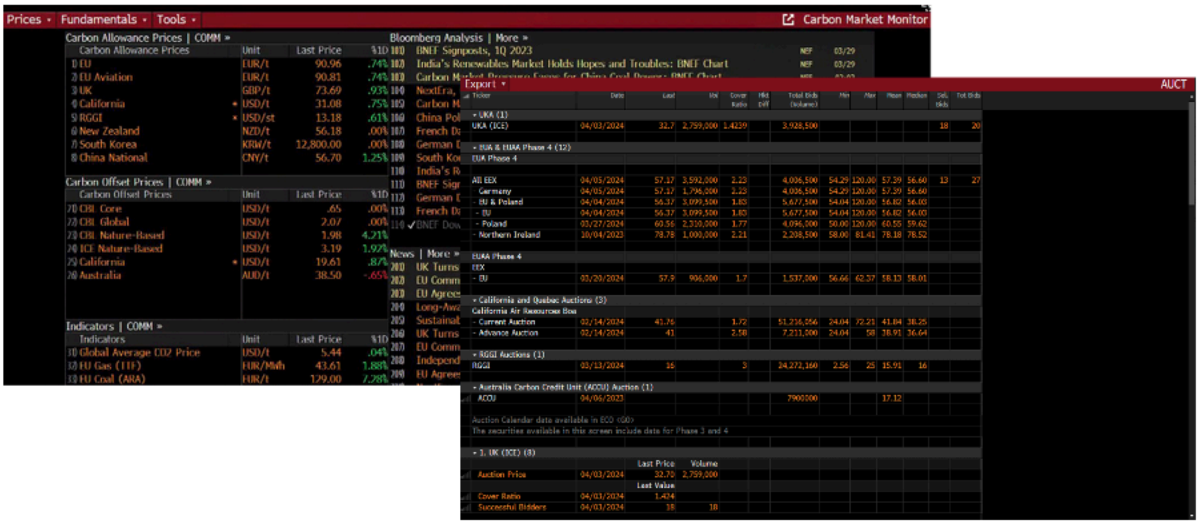


Figure 9 - CO2E and AUCT Screens on Bloomberg
(Source: *Bloomberg Terminal***)**

Researchers forecasting allowance prices have also used other types of data such as geo-spatial data such as temperature, precipitation, forest cover etc. from sources such as Google Earth Engine or GEE Community Catalog (by Dr. Sam Roy); Google Earth Engine is a multi-petabyte catalog of satellite imagery and geospatial datasets with planetary-scale analysis capabilities. Scientists, researchers, and developers use Earth Engine to detect changes, map trends, and quantify differences on the Earth's surface. Earth Engine is now available for commercial use, and remains free for academic and research use. GEE Community Catalog (awesome-gee-community-catalog) consists of community sourced geospatial datasets made available for use by the larger Google Earth Engine community and shared publicly as Earth Engine assets. IMF Climate Change Data is a website that offers climate change data and tools to support informed policy decisions and research while World Bank Open Data contains country-wise carbon emissions data in addition to other country level macroeconomic variables. Eionet, The European Environment Information and Observation Network (Eionet) is a partnership network of the European Environment Agency (EEA) and its 38 member and cooperating countries. EEA and Eionet gather and develop data, knowledge, and advice to policy makers about Europe's environment which is offered via the Industrial Emissions Portal. Finally, the well known data sources of YahooFinance, FRED and EIA provide information around stock indices and commodity futures, macroeconomic variables and energy price data respectively, which have sometimes been used as covariates in predictive modeling studies.

2.2 Feature Selection

(For full-detail tablation the selected features, data sources and preprocessing operations, refer *Appendix C*)

As discussed earlier in the literature review section as the second approach for carbon price forecasting (‘*feature-focussed machine learning approach*’), researchers have used features spanning *human factors* (also referred to as ‘socio-economic’ or ‘demographic’ factors), *institutional factors* (alternatively referred to as ‘regulatory’ or ‘policy’ factors), *market factors* (including ‘macroeconomic’ and ‘energy system’ covariates), *technological factors* (intended to study technological change and diffusion in decarbonization technologies) *and external factors* (such as ‘climate’, ‘conflict’ and ‘health’ related shocks) to predict the carbon price.

While other researchers have chosen or alluded to several explanatory variables, the authors of this study believed that it was institutional factors that govern how the markets are operating that could best provide explanations of how demand and supply interact and of how the price moves. This belief hence forms a prior for this study; This means that the authors have chosen several institutional or regulatory features as predictors such as the number of free allowances issued for each calendar year and at auctions, the transacted volumes, the emissions cap and how it is lowered, the sectoral coverage and the associated difficulty posed in abatement, the types of public-sector regimes governing the exchanges and their propensity to lean on permits regulation versus other forms of carbon taxes.

Additionally, from review of prior literature, the authors also concurred on the relevance of human factors in the context of the prediction problem. Three factors namely GDP/capita, Inflation and a Recession Indicator (which is set to 1 if GDP declines were noted for three prior quarters) were selected to this effect. From prior domain knowledge, the authors believed that while ‘GDP growth (%)’ and sharp rises therein was a good indicator of economic overheating, it was GDP per capita which was a more appropriate measure of how much of that overheating is passed through by the economy to the level of the household. Further the authors believed that owing to the inverse relation between CPI inflation and unemployment (as explained in macroeconomics by the ‘Phillips’ curve), that rising inflation could be a good proxy for economic overheating, and unemployment and thereby indicating human stresses. Since these factors originate in macroeconomics, they are classified as market factors in the data dictionary (Appendix C).

Finally, the authors believe that external factors do play an important role in the forecasting problem - the intuition for this is that when an extreme event occurs in the community, it drives up environmental consciousness. Consequently, the number of extreme weather events and the progression of the average global temperature anomaly in celsius (which is much studied by climate scientists), have been chosen as features for this study. Note that these features are defined differently in the context of different ETSs.

2.3 Data Preprocessing

(For full-detail on preprocessing operation for each vectors, refer *Appendix C*)

The target vector (allowance price), and feature vectors (described under section 2.2) used in this study are numerous. Since regulatory information about the carbon markets are mostly published in ad-hoc memos issued by the exchange, periodic bulletins or quarterly / annual summary reports, the data collection process was onerous, and involved scraping data from several sources to create structured datasets for modeling in the context of each ETS. Since Python APIs were not available for all data sources, GoogleSheets and python were used as the tools of choice for data collation - Python for gathering from apis, interpolation and imputation operations, and Google Sheets for assembling

data intuitively. The raw and processed datasets, and the preprocessing operations and tools built, in the context of this study may be found at -
https://github.com/pranavramkumar/wqu_capstone_5481/tree/main/data

Overarching data management operations (common to all ETS)

Allowance price imputation - There is sparsity in allowance prices sourced from Bloomberg or ICAP Allowance Price Explorer. This was corrected for by cubic spline interpolation between two available prices, followed by forward filling past the last available price, and back filling prior to the first available price. Since it is preferred and recommended to only retain trading days, since there is no price continuity outside trading windows, weekends were removed from the time-scale. For China, the Beijing and Shanghai allowance prices were averaged to construct a China average allowance price. (Not to be confused with the China National ETS, which is its own ETS focusing on the power sector).

Compliance period identifier - All ETSs have quantized time-periods within which a set of regulation prevails; These are called 'compliance periods' and are specified by the regulatory body. These have been collected for each ETS from their respective ICAP Carbon factsheets and thereafter encoded into a categorical representation.

Market data imputation - For census and macro data (such as GDP / inflation), sometimes values are available quarterly, while at other points only annually. This means that the value prevailing for a particular period is only reported by agencies at the end of the period. Consequently, the data must be backfilled for the entirety of the period. This is done to represent the activity during the period rather than focusing on the reaction to the data release, whilst realizing that policies may change upon data releases.

Regime type index construction - The purpose of constructing a regime type index ranging from 0 for right wing regimes to 1 for leftist regimes, is to determine how explanatory regime types and their policies can be on carbon pricing. Since there can be multiple member states under an ETS (such as the 10 states of RGGI, or the 27 member states within the EU), we first consider the set of member states to study, then identify the governor or head of state, their political party and its leanings. Next, we encode this leaning into a decimal between 0-1, and finally we weight each regime type index by the population of the respective member state to create a consolidated regime type index. For RGGI, the current 10 member states were considered for constructing the index, while for the EU, the regimes in France, Germany, Poland, EU Parliament, EU Commission, EU Council, Netherlands and Turkey were considered based on having the largest influence on EU and EU-ETS policy.

Merging recent data - Since macro or climate information tends to be reported after validation and on a lag, end of year 2023 data may not be available in many circumstances on larger databases such as US Census or WorldBank Opendata. However, smaller data companies or agencies may present end of 2023 data based on statistical methods or actuals. In such cases, some vectors have been prepared using data from one source until 2022, and from another source for 2023. Examples are state population data for RGGI from USA Facts (census aggregator) until 2022, and macrotrends for 2023, or state GDPs from USA Facts (census aggregator) until 2022, and bea for 2023.

Constructing a recession indicator - Since, we would like to evaluate the impact on price movements within recessionary periods, we construct a boolean recession indicator from the GDP data collected by checking if there was a continuous three quarter decline in GDP.

Constructing the technological diffusion indicator - As highlighted in researchers in prior carbon price forecasting or environmental economics literature, technological rebound has the largest effect on the carbon allotment price, since it marks a long-term shift in decarbonization technologies (albeit industrial or nature-based technologies). However, finding indicators to encode this information into the prediction problem has been a challenge. The authors have proposed an indicator, (despite understanding its limitations), towards capturing technological diffusion information. Cumulative Voluntary Carbon Market (VCM) issuances minus retirements for the two largest VCM credit originating sectors (energy and nature-based) has been used as the signal. This data was sourced from Quantum Commodity Intelligence data portal and scrubbed from the data dashboards.

Constructing the weighted average abatement difficulty indicator - Sectoral coverage for each ETS was studied on an annual basis, and a simple dictionary, encoded based on expert judgment and aware of sectoral decarbonization trajectories (shown below) was used to derive the average abatement difficulty score for the year.

Sector	Score
Fugitive Emissions	9
Agriculture	8
Maritime	7
Industry	6
Transport	5
Power	4
Waste	3
Buildings	2
Domestic Aviation	1

Table 4 - Dictionary of Sectoral Abatement Difficulty

Sourcing the average global temperature anomaly - The global temperature anomaly, measured by NOAA (since 1850), and which combines both land and ocean temperatures is sourced as-is from the NOAA data source, and used across all exchanges as a climatic indicator, because it is seen as the benchmark and is the main reference when making global warming claims. It is also a quantity that is forecasted into the future under climate scenarios.

Estimations - In the context of the Beijing and Shanghai pilot ETS, there is not sufficient data available around transacted volumes, or allowance allocation or auctioned allowance, or the emissions cap. However, government memos from the bureaus of ecology and environment provide cap reduction rules, daily volumes in the context of Beijing ETS, and other information around auctioned share and participating entities; This means that some unavailable information needs to be estimated using linear-regression fits, or annual data downsampled to a quarterly basis using representative ratios and splits from the available years. Sometimes linear regression fits have also been used to derive exchange data for the years 2017 or 2023 when unreported or unavailable in the context of other exchanges. The list of pre-processing estimations can be found at -
https://github.com/pranavramkumar/wqu_capstone_5481/tree/main/data/preprocessing

RGGI specific

Scrubbing regulatory information - Information around new free allowances, new auctioned allowances and traded volumes were obtained from the RGGI Market Monitor Reports. Information on cap, and cap reduction was obtained directly from RGGI bulletins. This was manually obtained as there are no APIs to connect with RGGI data.

Identifying other concurrent carbon taxes - In the context of the RGGI states, the number of prevailing carbon taxes at the state level is used as an institutional factor; It was obtained directly based on the published state statutes and regulations. This was manually obtained as there are no automated sources for this information.

Constructing the extreme climate indicator - The number of extreme climate events during the period from NOAA NWS (for the eastern seaboard states), was chosen as the extreme climate indicator in the context of RGGI. This was again scrapped from the portal directly, as there is no database around extreme event attribution.

California specific

Scrubbing regulatory information - For California, information around allocated allowances and auction results, and emissions caps were obtained from the respective CARB profile. Transacted volumes were obtained from the CARB Market Transfers Summary Report. There are no APIs in the context of California either, and hence the data was obtained manually.

Identifying other concurrent carbon taxes - In the context of California, the number of other prevailing carbon taxes in California was chosen and obtained from the CARB FAQ section, and from C2es.org; Once again, while this is an important explanatory factor, there is no clear structural source for policy data. This was hence obtained manually.

Constructing the extreme climate indicator - The number of extreme climate events (billion dollar events) from NOAA was considered as the external factor in the context of California. Once again since there is no single standardized source around extreme event attribution, this information was obtained manually.

EU ETS specific

Scrubbing regulatory information - Exchange information around new free allocation, auctioned allowances and daily traded volume, were obtained from the European Commission EU-ETS portal and from Bloomberg tickers on the CO2E screen respectively. As it can be observed, since much of the allocation, auctioning and volume information needed to be manually obtained from the exchange reports, the quality of compliance carbon markets information available on Bloomberg is not equally adequate across all markets particularly for the spot instruments. (Bloomberg does have rich information around live auction prices across all ~48 carbon exchanges globally during the progress of the auctions, and other providers such as Quantum Commodity Intelligence does provide sufficient news and transaction settling price information from the voluntary carbon markets).

Identifying other concurrent carbon taxes - Since sufficient detail is not available about prevailing carbon taxes in all 27 member states of the EU-ETS, and governmental tax revenue documents across each member state may be too many to unpack for a single indicator; Hence instead, 'Revenues from Environmental Taxes' across all EU member states from Eurostat was used as a proxy, and 2023 data was projected from prior years using a linear-model.

Constructing the extreme climate indicator - In the EU-ETS context, purely from a data availability perspective, 'Economic losses from extreme weather events' was used as a proxy for the number of extreme weather events. Once again this was obtained entirely manually with 2023 information estimated using a linear model fit.

Korea specific

Scrubbing regulatory information - Regulatory information around free and auctioned allowances are available from the annual K-ETS Summary reports from GIR. Data around trade volumes are available quarterly until 2020, has been downsampled using benchmarked quarterly trading activity from annual data for 2021 and 2022, and has been projected to grow at a 3% rate (based on prior year transaction growth rate) for 2023.

Identifying other concurrent carbon taxes - For Korea, the magnitude of Transport, Energy and Environmental taxes from Statista has been used as a proxy for the number of environmental taxes in Korea during the period, and 2023 data was estimated using a linear-model.

Constructing the extreme climate indicator - The extreme climate indicator chosen in the context of Korea was the sum of Heating Degree Days and Cooling Degree Days (from IEA) which are the days where the change in temperature upwards or downwards during the summer or winter exceeded a threshold level for the given month. This indicator is typically used in heat-health studies and is considered a relatively good proxy for extreme temperature effects. However the lack of an automated data source or api meant that it was sourced manually for this study and interpolated for 2021-2023 using a linear model.

China specific

Scrubbing regulatory information - Free allowances and new auctioned allowances data has been obtained from the benchmark auctioned share of 5% and 3% of the cap, for Beijing and Shanghai from the ICAP ETS Factsheets. Beijing Emissions Allowance (BEA) trade volumes for Beijing were obtained on a daily basis and aggregated annually from bjets.com.cn. For Shanghai, trading volume was obtained until 2020 from Lyu (2021) and estimated using a linear model for subsequent years.The total emissions cap for this model for China ETS (average of Beijing and Shanghai pilots)

Identifying other concurrent carbon taxes - In the context of China, there were six other prevailing carbon taxes during the period 2018-2021 as indicated by OECD. This was sourced manually from the paper. Since, other local sources have pointed to the existence or continuation of these taxes prior to 2018 and up to 2023, these were carried forward and back as the estimate for the entire period.

Constructing the extreme climate indicator - Beijing's air quality is a very highly studied variable in the context of China's international relations and climate reporting. While there are several structured api based weather sources around Beijing's air quality, it was sourced manually from NOAA for this study.

Finally, after constructing the structured datasets for each of the exchanges shown above, the target vector and a final selection of 14 vectors (X1-X14), were truncated into csv files for directly importing into python for the purpose of modeling. As it can be observed, the significant manual data sourcing for the purpose of feature extraction, means that model development from an industrial or commercial perspective can only be commissioned every 2-3 years or upon observing model drift. Auto-ML based methods are likely to be relatively unfeasible. Preliminary exploratory analysis performed on this data for the purpose of studying vector stationarity, collinearity, and data trends for the purpose of model selection is presented under section 4.1 and 4.2.

3. METHODOLOGY

This paper approaches the problem of carbon price forecasting using machine-learning and deep-learning based modeling in order to incorporate the benefits of non-linearity, lags and memory, different cross-validation approaches, and hyper parameter optimization. The paradigms used to make predictions, and their theoretical foundations are as described below.

3.1 CEEMDAN-LSTM

As viewed from the literature review on prior forecasting literature, there are three broad categories of models which have historically been used to forecast the carbon price namely - (i) statistical models (such as ARCH, ARIMA, GARCH and variants) (ii) feature-rich machine learning models and (iii) hybrid models comprising a form of signal decomposition (to delineate non-linearity and non-stationarity) and a machine learning based prediction model. CEEMDAN-LSTM which expands to Complete ensemble empirical mode decomposition with adaptive noise + Long Short Term Memory model falls into the third category of models described above.

What the first component of the hybrid-model i.e. CEEMDAN (discovered by Torres et al (2011)) does, is breakdown the predicted signal into a series of ‘modes’, ‘inherent mode functions’, or ‘eigenmodal components’. These modes are components which describe the periodicity and trend of the predicted variable, which may have recurring patterns over a variety of time-scales. In the context of the carbon allowance price forecasting problem, this can be interpreted as attributing today’s price shift to a period event which occurs every several years, as opposed to a more recent event. This process of mode decomposition performed under CEEMDAN, is one of several types of a more general practice in signal processing called ‘spectral decomposition’. The process of how a spectral decomposition algorithm iteratively breaks down a signal into patterns at different time-scales is shown in Figure 10.

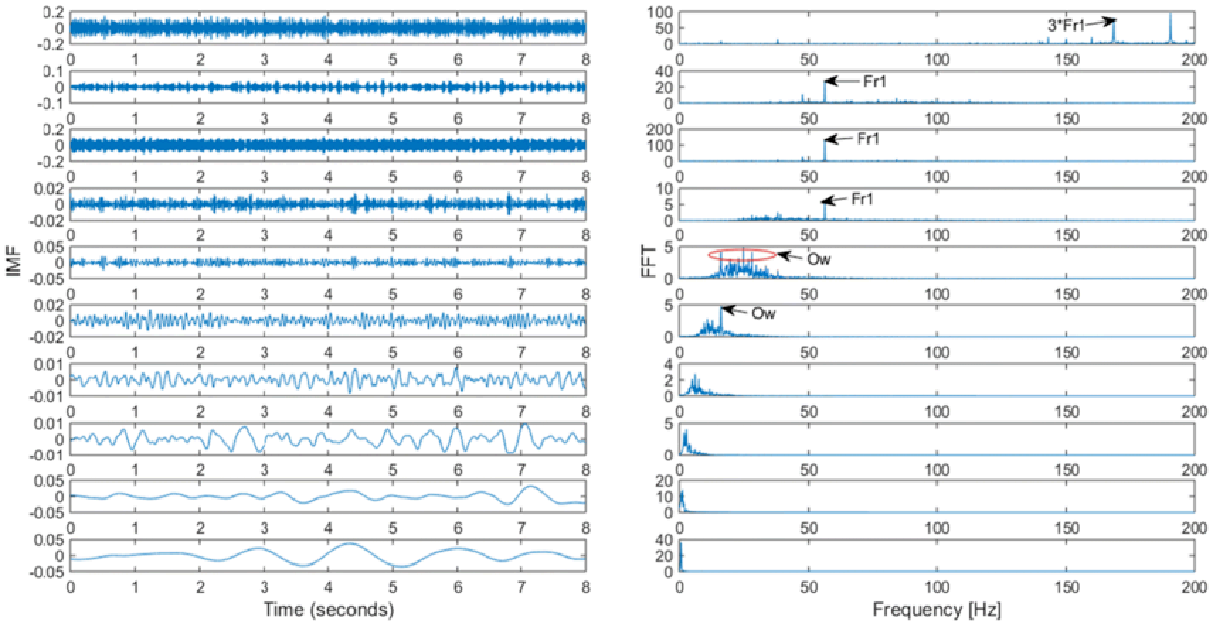


Figure 10 - Mode Decomposition performed by CEEMDAN algorithm, and their corresponding Fast Fourier Transform Spectra (Source: Kebabsa and Djebala (2020))

A list of potential spectral decomposition methods and their uniqueness is presented below in Table 5.

Acronym	Method	What it does	What makes it unique
EMD	Empirical mode decomposition	It is a method used to decompose nonlinear and non-stationary time-series data into a series of inherent mode functions (IMFs) with various fluctuation scales.	Foundational method for signal decomposition
EEMD	Ensemble empirical mode decomposition	Due to intermittent raw data source, the modes generated by EMD could become confused. Hence EEMD adds a Gaussian white noise term and performs repeated decomposition	Adds Gaussian white noise prior to signal decomposition
MEMD	Multivariate empirical mode decomposition	It is a form of spectral decomposition that extends 1 dimensional EMD for a multiple dimensional signals.	It is a multi-dimensional form of EMD apply to multichannel or multi output scenarios
VMD	Variational mode decomposition	VMD allows us to control the band-width of the decomposed input signal under WMD. It provides improvements over regular EMD such as no modal aliasing effect and sensitivity to noise	It can attain no modal aliasing effect and sensitivity to noise
Fast-EMD	Fast empirical mode decomposition	It is a method for decomposing multi-dimensional signals faster by concatenates multi-variate or multidimensional signals into a one-dimensional signal and uses various one-dimensional EMD algorithms to decompose it	Performs serialization of multi-dimensional input
WT	Wavelet transformation	Wavelet transforms are mathematical tools for analyzing data where features vary over different scales. For signals, features can be frequencies varying over time, transients, or slowly varying trends	A wavelet, unlike a sine wave, is a rapidly decaying, wave-like oscillation. This enables wavelets to represent data across multiple (shifted) scales.
SSA	Singular spectrum analysis	SSA is a nonparametric signal decomposition method, which addresses problems of finite sample length and noisiness of sampled time series not by fitting an assumed model to the available series, by using a data-adaptive basis set, instead of the fixed sine and cosine of the basis transform method.	A non-parametric approach to signal decomposition using a data-adaptive basis transformation
CEEMDAN	Complete ensemble empirical mode decomposition with adaptive noise	Residual white noise still affects the accuracy of EEMD. CEEMDAN addresses this reconstruction error by introducing adaptive white noise at each stage of the transform	Addresses this reconstruction error in EEMD by introducing adaptive white noise at each stage of the transform
ICEEMDAN	Improved complete ensemble empirical mode decomposition with adaptive noise	ICEEMDAN improves upon the benefits of CEEMDAN by reconstructing the useful signal from the Intrinsic Mode Functions (IMFs) using an improved adaptive shrinkage scheme related to interval extremum based on an interval thresholding (IT) function.	Improves the reconstruction of CEEMDAN approach using interval thresholding

Table 5 - List of spectral decomposition methods

The six steps within the CEEMDAN approach for spectral decomposition are as follows -

1. Gaussian white noise is added to the residual sequence to obtain an updated sequence $\overline{y_i}(t) = y(t) + \sigma n_i(t),\ i = 1, 2 .... N$

2. Empirical Mode Decomposition is performed on this updated sequence and the resulting N (8 in our case are averaged) model components are averaged to obtain the first IMF of the CEEMDAN approach, and the first residual component is calculated.

$$imf_1(t) = \frac{1}{N}\sum_{i=1}^{N} imf_{1i}(t) \text{ and } R_1(t) = \overline{y}_i(t) - imf'_1(t)$$

3. The adaptive white noise $\sigma E_1(n_i(t))$ is added to this new residual $R_1(t)$ and EMD is performed again to obtain $imf_2(t)$
4. This process is repeated continuously until all modes are obtained $imf_n(t)$ and no further EMD is possible
5. Finally the original sequence can be re-assembled from the eigenmodal components (imfs) with a residual trend component as below -
   $y(t) = imf(t) + R_{es}(t)$
6. After CEEMDAN has decomposed the residual series, in our case LSTM based forecasting is applied to each imf, and finally these forecasts are linearly combined to obtain the forecast sequence $\hat{y}$.

Theory about the LSTM Recurrent Neural Network developed by Hochreiter and Schmidhuber in 1997 is presented below. Like other RNNs, LSTMs can send information between the same or previous layers through interlinkages in their design (cells / neurons).

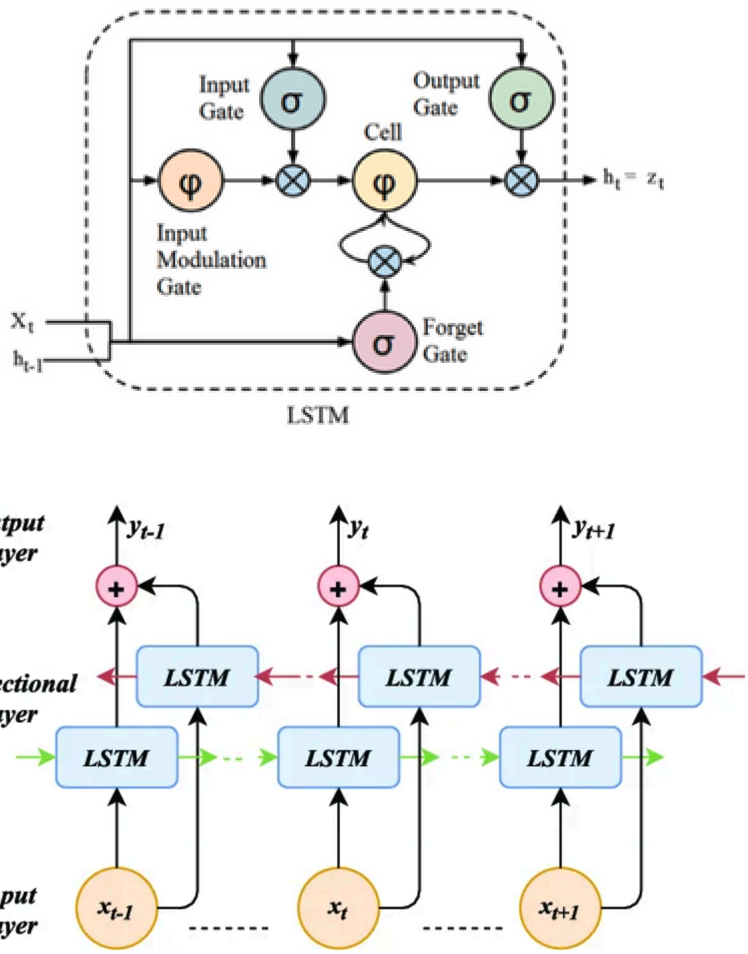


Figure 11 - Schematic of LSTM model
(Sources: Anishnama, Medium and Shahzad et al (2023))

Further the ability of LSTM networks to handle short term memory in addition to long-term memory improves its predictive performance. Each cell is a building block comprising of three types of gates each activated by a sigmoid function, which allows to control the amount of information entering or leaving the cell. The gates are -

1. Forget gate: outputs a value between 0 and 1 denoting that the data entering the cell must be completely retained or completely ignored, with fractional values denoting an intermediate approach.
2. Input gate: which determines the amount of new data entering the neuron to actually save in the cell. The workings of this is controlled by a sigmoid and a tanh activation layer.
3. Output gate: determines what the cell outputs, based on the state of the cell and the most recent value which enters the cell.

The LSTM equations (popularly used to denote the model) are -

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + b_f)$$
$$i_t = \sigma(W_{ix}x_t + W_{ih}h_{t-1} + b_i)$$
$$\sigma_t = \sigma(W_{ox}x_t + W_{oh}h_{t-1} + b_o)$$
$$c_t = f_t \odot c_{t-1} \odot tanh(W_{cx}x_t + W_{ch}h_{t-1} + b_c)$$
$$h_t = o_t \odot tanh(c_t)$$

where,
c_t - denotes the value in the memory cell, f_t - denotes the forget gate, i_t denotes the input gate, and o_t denotes the output gate. $\odot$ denotes element wise product operations, x_t denotes the value of the input vector and h_t denotes the value of the hidden state vector of the LSTM at time t.
W and b are the weights matrix, and bias terms which needs to be learnt in model training;
σ denotes the sigmoid activation function, and tanh denotes the hyperbolic tangent activation function

3.2 Boosting and XG Boost

Gradient Boosting is a form of supervised machine learning (training using labelled data), where predictions are obtained by a set of weak learners (typically decision trees) combined in a stage-wise fashion to reduce the residual error by minimizing a cost function using a gradient descent.

XGBoost (created by Chen et al, 2016) is a special case of gradient boosting with an optimized boosting process - which performs second order expansion of the loss function, controls model complexity through regularization to avoid overfitting and splits nodes in a manner so as to better capture non-linear relationships between features in model training.

XGBoost is designed to handle both classification and regression problems, and its ability to handle missing values and work with sparse data makes it versatile for a wide range of applications. The algorithm is renowned for its capability to handle large-scale datasets, parallelize computations, and prevent overfitting through advanced regularization methods like L1 and L2 regularization. XGBoost's combination of boosting and regularization is what makes it very successful.

The hyperparameters of XG Boost are -

- *n_estimators*: (The number of boosting rounds or weak learners)
- *learning_rate*: which controls the contribution of each weak learner to the final prediction
- *max_depth*: The maximum depth of each decision tree weak learner
- *min_child_weight*: The minimum sum of instance weight (hessian) needed in a child to prevent overfitting

- *subsample*: The fraction of samples to be used for fitting the individual weak learners (between 0 and 1, introduces randomness)
- *colsample_bytree*: The fraction of features to be used for fitting each decision tree. It introduces additional randomness of each tree and can prevent overfitting
- *gamma*: The minimum loss reduction required to make a further partition on a leaf node of the tree. It helps control tree growth
- *alpha*: L1 regularization term on leaf weights. It can be used to add L1 regularization to prevent overfitting
- *lambda*: L2 regularization term on leaf weights. It can be used to add L2 regularization to prevent overfitting
- *scale_pos_weight*: Controls the balance of positive and negative weights (for imbalanced datasets)
- *objective*: The loss function to be minimized ('reg:squared error', 'binary:logistic', and 'multi:softmax')
- *eval_metric*: The evaluation metric used during training to measure performance on the validation set
- *random_state*: seed for randomizer

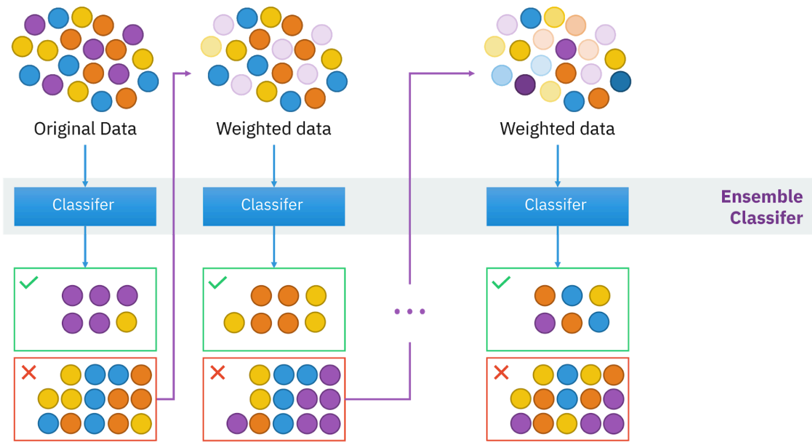


Figure 12 - Schematic of Gradient Boosting and Extreme Gradient Boosting (XGBoost) model (Source: neptune.ai)

3.3 Transformers

Transformers are a class of encoder-decoder models (sequential neural networks), which are specifically trained for sequence to sequence (Seq2Seq) tasks such as natural language processing, time-series models similar to RNNs such as GRU or LSTM, except that they have a mechanism called “attention”. In the context of transformers each observation in the input data stream is represented as a positional encoding, i.e has value by virtue of its position in the sequence (such as words in a sentence), which introduces similar capabilities as auto-regressive or recurrent neural-net models to encode memory and sequential information. The “attention” mechanism involves mapping inputs to latent feature vectors, which is then further mapped to output space.

A transformer model consists of a number of ‘transformer layers’ each of which comprises a multi-head self attention (MHSA) mechanism and a fully connected feed forward network (FFN) mechanism as shown below. In a time-series context, the model is trained by comparing encoder output to the next time-step's value and iteratively trying to reduce the error between the two.

The field of carbon pricing has not used transformer models heavily with the exception of some Seq2Seq models; This paper hence decided to adopt a simple transformer architecture to see how the model performs in the forecasting problem.

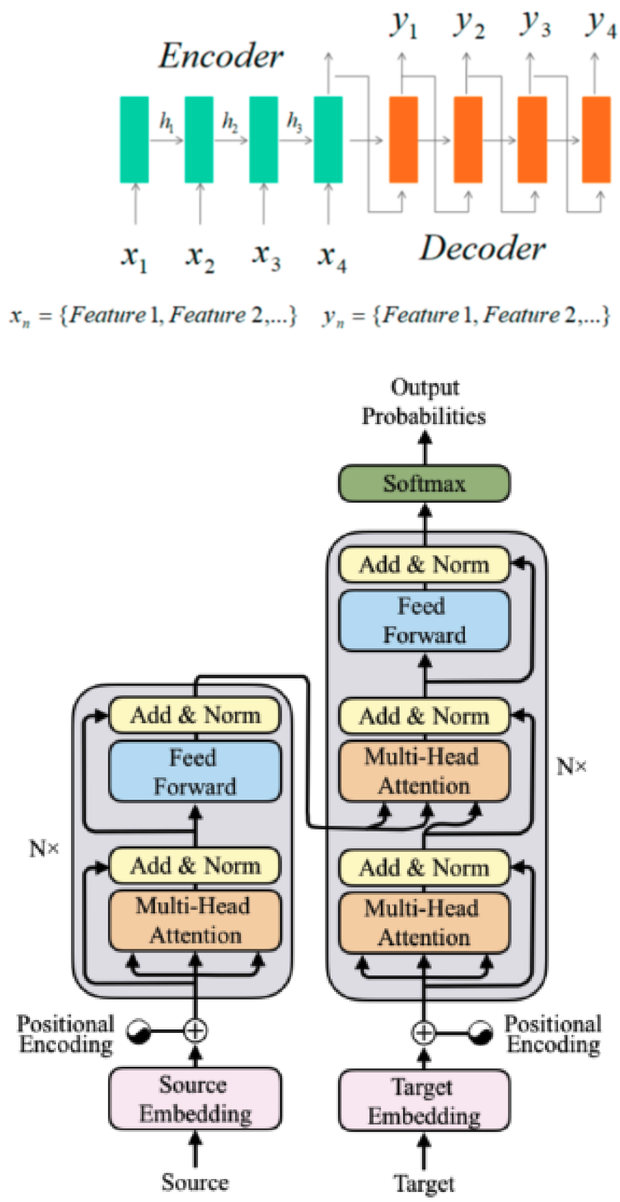


Figure 13 - Schematic of Transformer Model
(Source: Zhang et al (2021) and Hwang et al (2019))

3.4 Train-Test splits used

For the XGBoost and Transformer models, the train, test and validation (unseen) splits were chosen as 70%, 15% and 15% of the data respectively. Consequently the out-of-sample forecasts seen later in section 4 on the unseen data was for a period of ~ 1 year (all of 2023), which roughly represents about 15% of the data.

In the interest of speed of execution, the train-test-validation splits for the CEEMDAN-LSTM model were chosen differently; The validation split consisted of only 30 observations, while the rest of the data was split 80%-20% between the training and test splits. Out of sample performance on the unseen split is hence reported for only a period of about 30 days.

3.5 Loss functions used

The models were initiated to compute and report the following loss functions in model training (in-sample) and testing (out-of-sample). Out of sample loss functions are reported in the results section and on Appendix E.

Mean Absolute Error(MAE)	$\frac{1}{n}\sum_{t=1}^{n}\left\lvert\widehat{y_t} - y_t\right\rvert$
Root Mean Squared Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{t=1}^{n}(\widehat{y_t} - y_t)^2}$
Mean Absolute Percentage Error (MAPE)	$\frac{1}{n}\sum_{t=1}^{n}\left\lvert\frac{y_t - \widehat{y_t}}{y_t}\right\rvert$

Table 6 - Loss Functions

4. RESULTS

This section *first*, uses exploratory data analysis to answer certain inquiries of the carbon allowance prices, and carbon allowances as an asset class. *Second*, it expressly observes certain diagnostic plots of the chosen feature space used for forecasting and plots of the target vector in order to concur on the rationale for the use of models which account for nonlinearity, non-stationarity and multiple time-scales in the data. *Finally*, this section reports the out-of-sample performance from the models trained to forecast carbon allowance prices in the context of each of the five ETS in-scope of this study, and comments on approaches to improve the model performance.

4.1 Exploratory Analysis

An exploratory analysis was performed using market data to validate certain priors which motivated this study and also guides the construction of a commercial ETF using the carbon allotment price forecasts developed in this paper.

First, a plot of trailing 5 year allowance prices of the five (six in actuality, considering China includes Beijing and Shanghai ETS) ETSs in scope of this study confirms the prior that carbon spot allowance price movements are inherently 'jumpy' due to the large impact of regulatory changes, human factors or extreme events in guiding price moves.

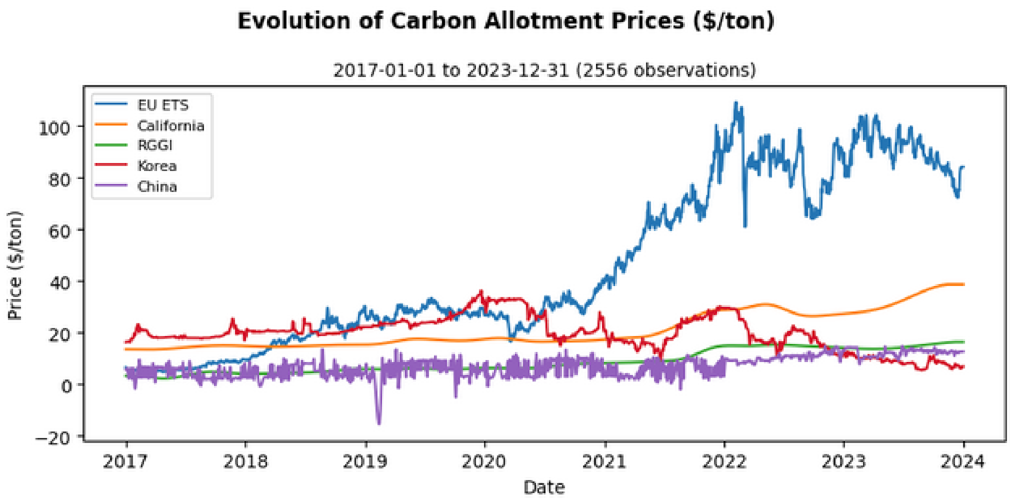


Figure 14 - Evolution of Carbon Allotment Prices (2017-01-01 to 2023-12-31)

Second, choosing to securitize an ETF product using carbon allowances was premised on the intuition that that carbon allowances as an asset class were uncorrelated with any other kind of asset class (currency, commodity, stock index, metals, emerging markets bond or equity index etc..). Constructing an equi weighted carbon index using the five ETS prices, and plotting a correlation matrix versus other equity (S&P 500), commodity (DBC), metal (gold), emerging market bond (EMB) or emerging market equity (EEM) indices confirmed the intuition into fact, that carbon prices were in fact uncorrelated.

	Carbon Index	DBC	EEM	EMB	GLD	^GSPC
Carbon Index	1.00	0.06	0.06	0.03	0.03	0.04
DBC	0.06	1.00	0.38	0.28	0.26	0.36
EEM	0.06	0.38	1.00	0.59	0.19	0.77
EMB	0.03	0.28	0.59	1.00	0.30	0.59
GLD	0.03	0.26	0.19	0.30	1.00	0.08
^GSPC	0.04	0.36	0.77	0.59	0.08	1.00

Figure 15 - Correlation of a Carbon Index with Different Asset Classes

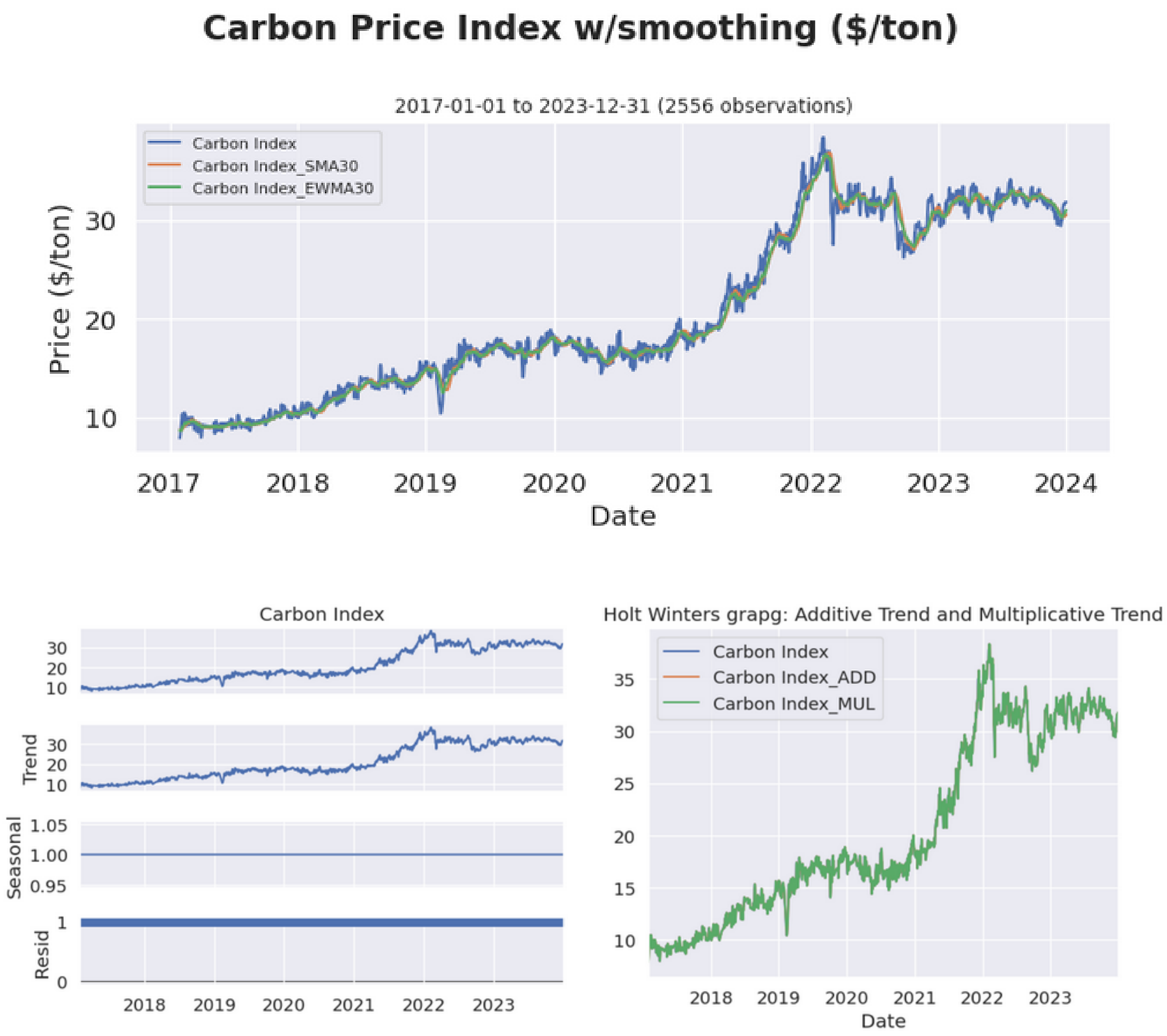


Figure 16 - Constructing a Smoothed Carbon Price Index

Finally, the authors believed that since allowance prices demonstrated significant volatility and jumps, that they would not necessarily present themselves innately as an investable asset class as a single instrument. The exploratory analysis sought to confirm if it was possible to construct a ‘smooth signal’ through a combination of the five ETS price tickers in some portfolio combination. IN order to validate this, an equi weighted price index of the five ETS prices was smoothed using three approaches - Simple Moving Average, Exponential Moving Average and Holt Winters seasonal decomposition. The figure above shows that it is possible to create a carbon index through portfolio combination or optimization that has smooth and continuous price history.

4.2 Diagnostics

Diagnostic plots of the feature and target vectors were construed for each ETS and observed to see that the explanatory vectors were inherently nonlinear, nonstationary and had different temporal patterns, which warranted the application of models such as modal decomposition hybrid models, or machine learning models in forecasting.

The figure below shows the diagnostic plots in the context of the Regional Greenhouse Gas initiative.

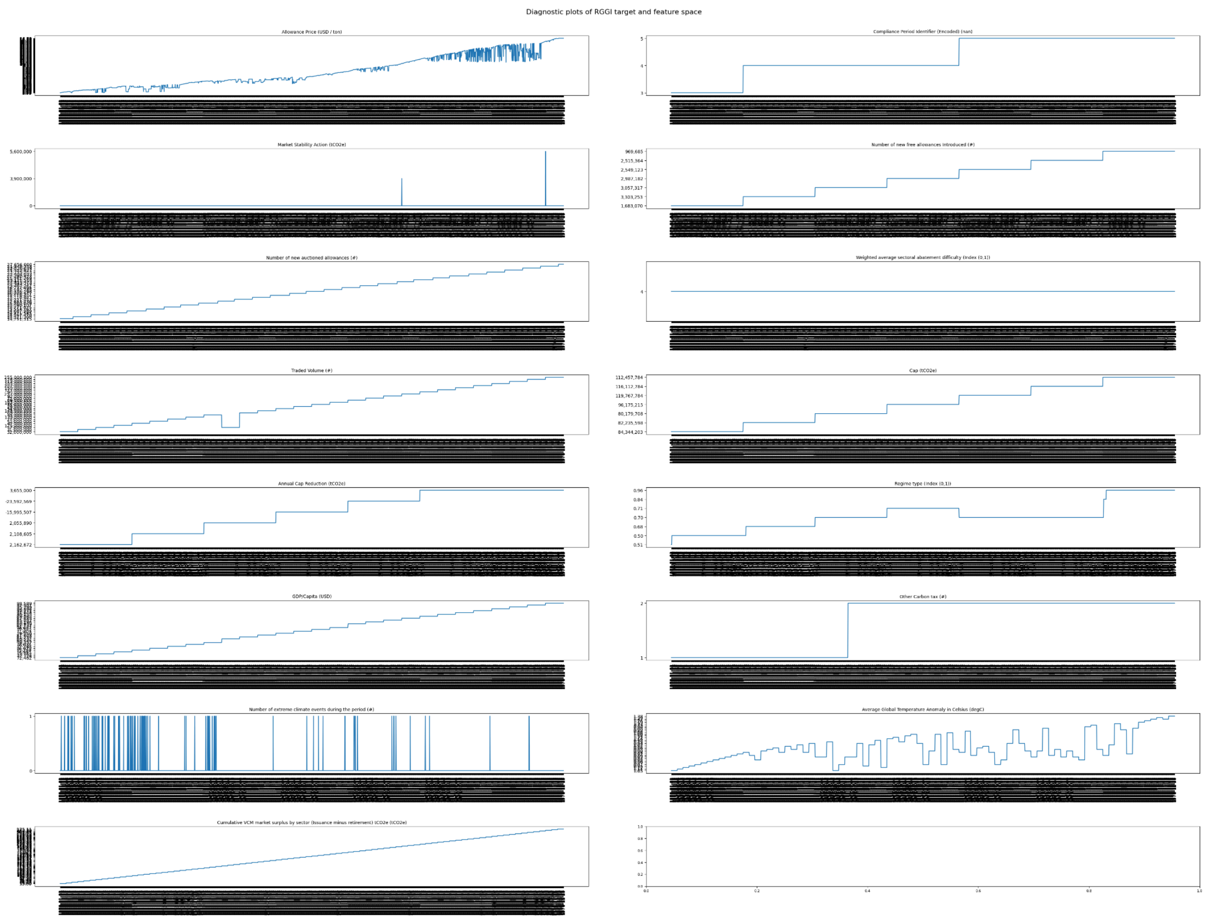
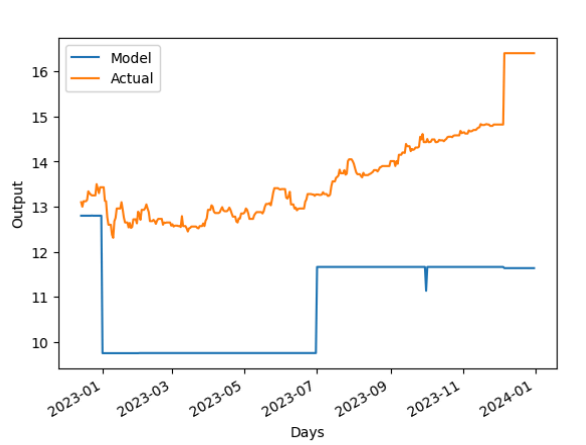
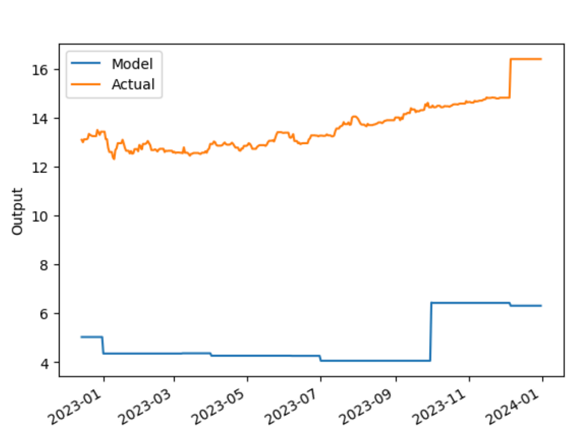
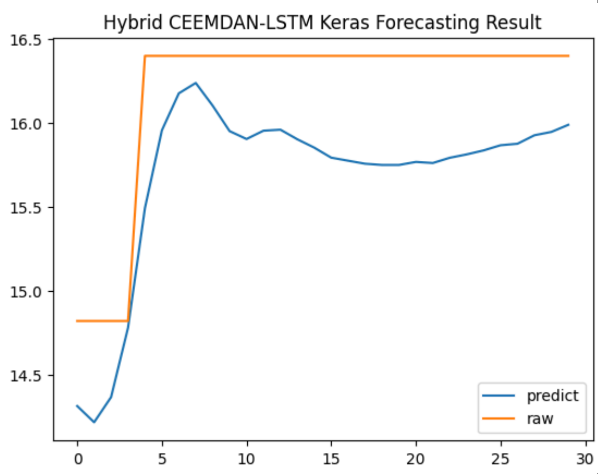


Figure 17 - Diagnostic plots of the feature space - RGGI

4.3 Forecasting

This section is dedicated to providing the details of the predictive modeling performed, and comparing the results of the XGBoost, Transformer and CEEMDAN-LSTM models in forecasting the carbon allowance prices, on unseen samples of data.

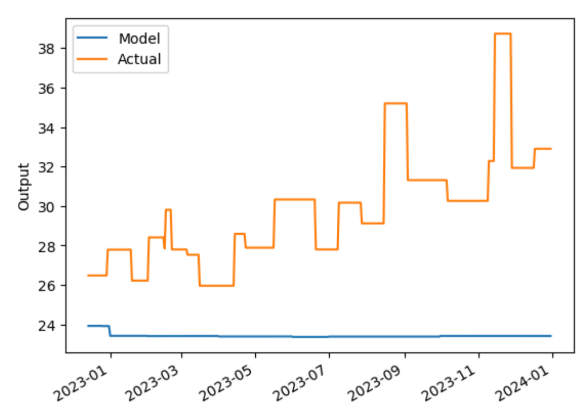
	
XG Boost	**Transformer**



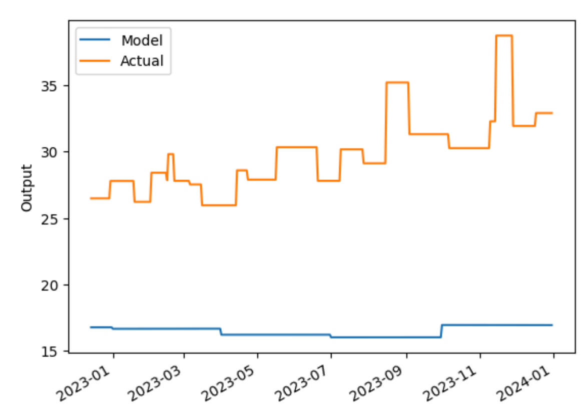
CEEMDAN-LSTM

Best Model: CEEMDAN-LSTM
Best Model MAE: 0.5036
Best Model RMSE: 0.5295
Best Model MAPE: 0.0310

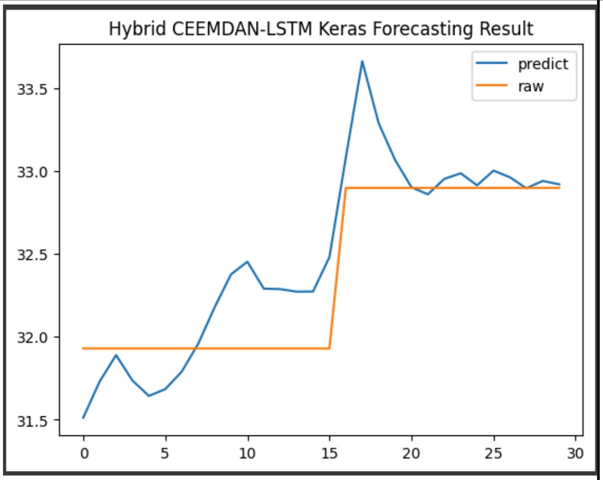
Figure 18 - Out of Sample Predictive Accuracy - RGGI



XG Boost



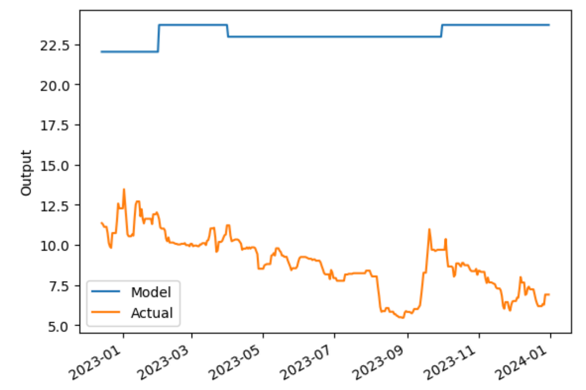
Transformer



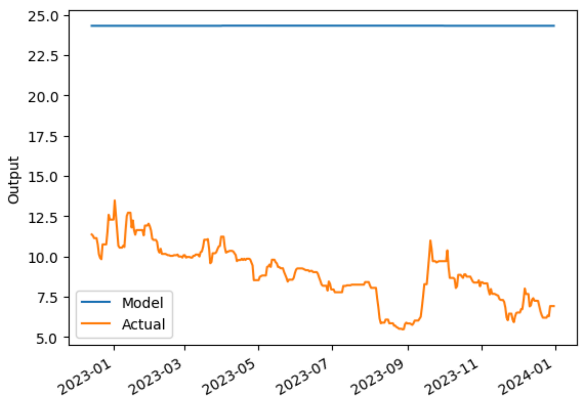
CEEMDAN-LSTM

Best Model: CEEMDAN-LSTM
Best Model MAE: 0.2579
Best Model RMSE: 0.3090
Best Model MAPE: 0.0080

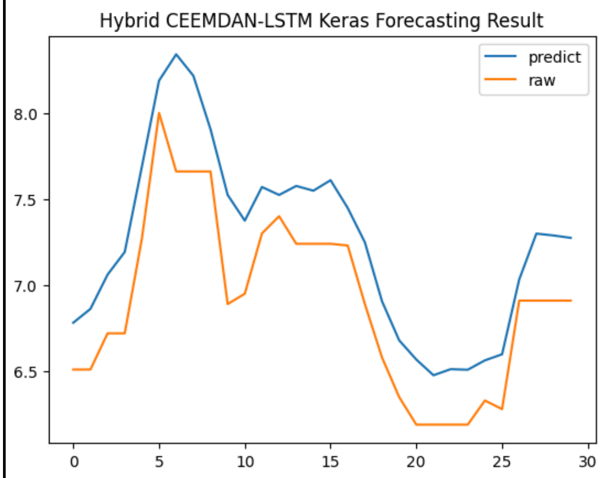
Figure 19 - Out of Sample Predictive Accuracy - California ETS



XG Boost



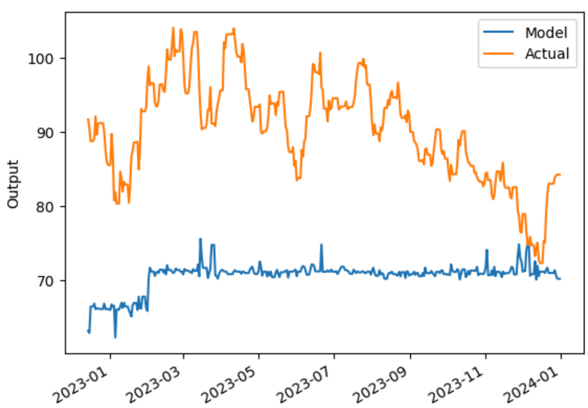
Transformer



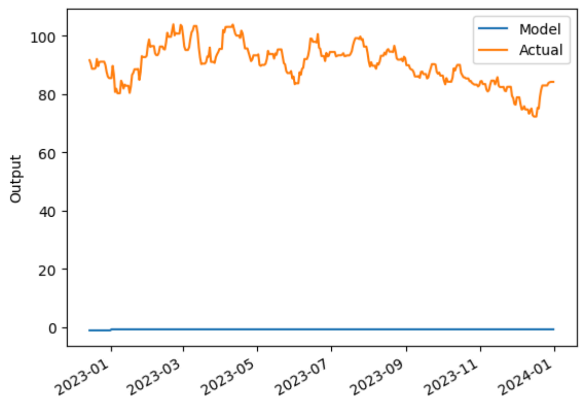
CEEMDAN-LSTM

Best Model: CEEMDAN-LSTM
Best Model MAE:
Best Model RMSE:
Best Model MAPE:

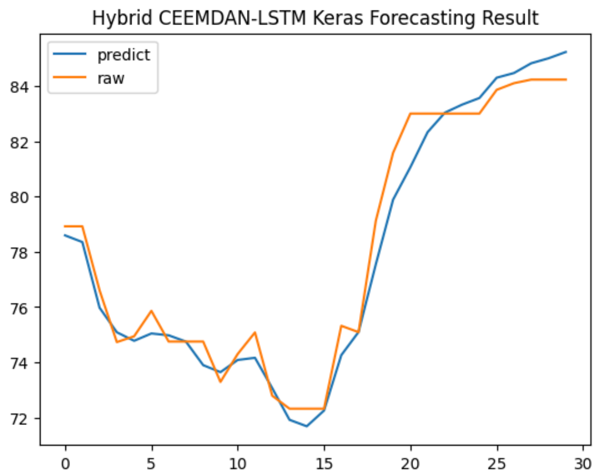
Figure 20 - Predictive Accuracy - Korea ETS



XG Boost



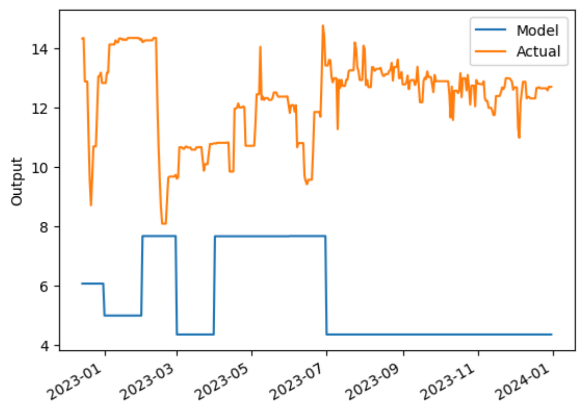
Transformer



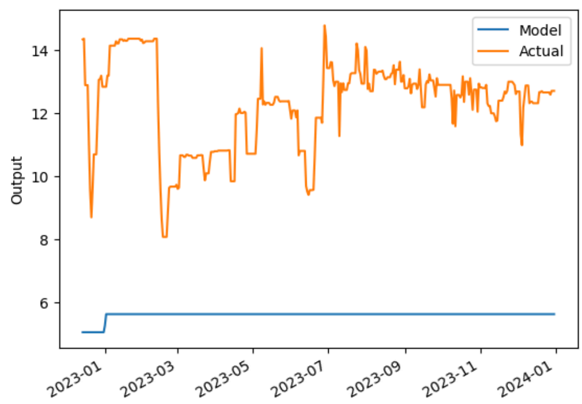
CEEMDAN-LSTM

Best Model: CEEMDAN-LSTM
Best Model MAE: 0.5948
Best Model RMSE: 0.7632
Best Model MAPE: 0.0075

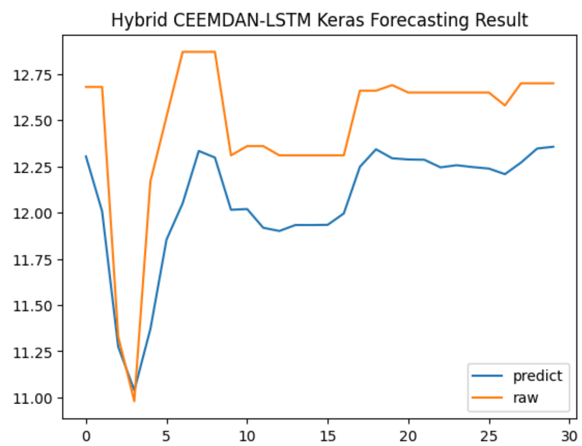
Figure 21 - Predictive Accuracy - EU ETS



XG Boost



Transformer



CEEMDAN-LSTM

Best Model: CEEMDAN-LSTM
Best Model MAE: 0.5948
Best Model RMSE: 0.7632
Best Model MAPE: 0.0075

Figure 22 - Predictive Accuracy - China ETS (Beijing + Shanghai)

Observing the model estimation results shows that CEEMDAN-LSTM is unanimously the best performing model across all ETSs. The XGBoost and Transformer models, on average, significantly underestimate carbon allowance prices. We may interpret the above as evidence of the following -

1) Feature space:
The underperformance of XGBoost (given its very optimized nature of modeling, and given the fact that hyperparameters were fully optimized using GridSearch), conveys that the feature set does not bring in sufficient information in the context of the period 2017-2023, for the prediction problem at hand i.e. a feature space predominantly comprised of institutional and regulatory factors is not sufficient for forecasting allowance prices. Other factors such as human, market, technological and external factors could be carefully chosen and moreso in the context of each specific ETS. Just as human demographic information has been proxied using energy volatilities, other proxies may be considered to represent or encode other hard-to-obtain information.

2) Data pipeline and hardware:
The lack of a well structured data pipeline for the compliance markets and the need to manually construct the datasets using parser layers is onerous and a very big barrier to training new models, more suited for larger data spaces. Efforts to enhance the data pipeline or leaning on paid services which are making these efforts, could be a useful consideration for future price prediction efforts. Further, when it came to experimenting with Transformer architecture, we were limited by the hardware available. For a 9 feature input, we could not have more than 18 layers before we experienced a crash due to RAM usage.

3) Non-Sequential information:
The under-performance of the transformer model brings into question the sequential nature of carbon price data or the kinds of regulatory information used in the prediction problem. Positional encoding of the data for the transformer may not be accurate with only data from 2017-2023 due to the multi-temporal nature of the carbon price signal. Further testing with more robust practices such as data encoding, standard scaling, different transformer architectures, or combining transformer with GARCH models (such as MT-GARCH) with supplemented volatilities in the feature space, could reveal if transformers form a good model class for this prediction problem.

4) Train-test splits:
In this study, the authors were only able to consider a naive train-test split mechanism, which may be inadequate. This may be addressed using k-fold cross validation (which addresses one-sidedness of test-set selection), walk-forward validations (which does away with the effects of anomalous periods such as COVID-19 through window-ed train and test splits), or combinatorial purged cross validations (which purges and embargos data prior to test splits, in order to prevent information leakage from the training set).

5) COVID-19 economics:
The period 2019-2021 was very anomalous for global emissions due to lockdowns, lower than expected transport emissions, and industry consolidations. Regulatory and institutional factors may be inadequate to explain carbon price movements during this period, and supplementing the feature space with human factors such as household incomes or daily covid-19 new cases could be a meaningful indicators of allowance prices during countercyclical macroeconomic periods.

Solving for the above in subsequent model development, or re-calibration could serve to improve model performance.

5. DISCUSSIONS

This section presents discussions regarding the challenges expected when transitioning from making carbon price predictions using deep learning, towards being able to use the forecasts to create a well-functioning carbon ETF. The discussions are organized into four topics namely -

- *Environmental ETF Design*, which serves as a literature review on the properties and features of an exchange traded fund and a competitor analysis of other contemporary carbon ETFs
- *Environmental ETF Construction*, which discusses several additional issues to be considered or modeled into the prediction problem for robust fund-emulation and backtesting
- *Climate Scenarios*, which serves as a literature review of standard scenarios used by the global industrial and financial sector, through scientific consensus to forecast climatic variables and emissions into the long-term
- *Long-term forecasts*, how to rely upon statistical techniques, climate models or other long-term macroeconomic forecasting models to generate forecasts of the feature space out to 2030 or 2050 to back-test the alignment of fund environmental impact with global environmental targets

5.1 Environmental ETF Design

An environmental or carbon ETF, has to follow similar protocols for issuance, registration and regulation as other exchange traded funds. In short, the ETF prospectus and design must be filed with the appropriate governing body, a trust must be created to hold the stock shares as the underlying security, and shares created in bundles must be sold on the open market. These shares operate like normal stock shares on the open market, but are open to different market forces due to the underlying securities within the ETF. Further details on the ETF planning process are described below.

First, the issuer or proponent of the ETF must plan the overarching goal and operations of the ETF; The overarching goal could be index tracking and providing comparative financial returns or attaining environmental impact (aligned to 2030, or 2050 goals) together with creating an investment vehicle. Planning the operations of the ETF involve identification of the parties namely the *fund advisor* who plans the fund composition and asset weightage, *calculation agent* who computes and maintains any indices, *fund administrator* who buys and sells underlyings into an entity based on the fund's mandate*, trustee* who maintains custody of the securities owned by the fund, and lead *portfolio manager*, a key person who directs the fund. It also involves planning and executing the legal entity structure for the fund. A popular structure is 'fund + subsidiary', where the fund entity raises capital for investment, and the subsidiary invests a certain fraction of the amount raised (~25-50%). An alternative structure if the fund is created as part of a large asset management company is 'fund + sub-fund / sleeve', where the multiple funds or a series of funds can all be administered as part of the same sub-fund (typically a subsidiary of the fund holding company).

Second, the ETF must be registered with the lead securities regulator in the country of issuance (E.g. Securities and Exchange Commission (SEC) in the United States), and since the ETF has to be traded, a ticker symbol will have to be issued by the exchange when the ETF Prospectus is registered. An ETF Prospectus will have to contain all relevant information pertaining to the fund in full detail such as - Issuer or Proponent, Portfolio Manager, Market Cap / Net Assets, Thematic Interest, Launch Date, Ticker, Exchange, Goal, Markets, Underlyings, Investment Specifications, Currency, Legal Structure, Domicile, Applicable Law, Regulators, Investor Qualification, Minimum Investment, Return, Risks, Fees and expenses. Refer Appendix C to see condensed prospectus information for nine contemporary Carbon ETFs in the market)

Finally, the fund manager will have to disclose specific investment risks from the specific approach taken to generate and unwind exposures in the investment process; The fund will need to report performance aligned to fund performance reporting standards appropriate for the issuing market. The fund prospectus will also need to divulge all investment risks such as loss of principal, non-linearity, leverage, market, credit and liquidity risks, calculation and tracking risks etc.

For a tabular summary of several contemporary Carbon ETFs, refer *Appendix F*.

5.2 Environmental ETF Construction

Having observed the features, legal structures and risks of several contemporary carbon ETFs under 5.1, we anticipate the following technicalities in progressing from a well functioning deep-learning based forecasting solution to a commercial ETF -

Exposures
The study above used a feature set to forecast carbon allowance 'spot' prices; However seeking exposure to the carbon price as a fund may require on to participate in the equivalent futures market for two reasons - spot allowances being reserved for allocations to companies by regulatory bodies limiting liquidity available for purchase, and the fact that futures price in regulatory changes differently from spot instruments, and it could serve as a means of avoiding negative price impact from anticipated upcoming regulatory shifts.

Vintages
It would not be right to assume that carbon allotments are not fungible over the entire period of study (2017-2023). Regulatory bodies issue new free allowances as part of the annual allocations, or under auctions where some issued allowances are for the current year, while some are for future years (with a specific vintage marked on the issued allowance) (refer CARB). This means two things - buyers or companies of carbon allowances may not purchase older vintages in circulation for they may not be able to surrender them to offset emissions under the current compliance periods, and allowances purchased ahead of schedule for future periods will need to necessarily banked or surrendered by the marked vintage date, thereby capping the potential rise in carbon prices from trading the allowance in the interim. This behavior needs to be considered in detail into the modeling problem if possible to capture inter-temporal and inter-year effects in allowance allocation and trading.

Banking of Allowances
In this study, the allowance allocation and auctioned allowance information used as features are those 'net' of banking activity i.e. the numbers reported as incremental additions to the allowance pool by exchanges net of any allowances surrendered by emitting companies, banked by environmental groups or removed from circulation by exchanges. Explicitly incorporating this as its own feature may improve the prediction performance and also help to capture the difference in time-intervals between actual new allowance issuances and the time to bank the credits.

Anticipating Policy
Currently regulatory action, macro-economic numbers, and climate event information is captured into the feature space ex-post (i.e. after it is reported). An experienced forecaster or commodity markets practitioner could argue that this is already too late for the information to be considered and make a difference to the prediction problem. There is hence the need to build leading indicators - either through proxying variables, or by applying adjustments or projections by observing event horizon information (in the regulatory or climate domain) or observing news sentiment and calibrating the latency between news sentiment and policy change.

Localized Effects
The study in the paper only considered broad climate variables as explanatory factors i.e. the number of total climate events across the entire state or country, or the global temperature anomaly. While this does hold some explanatory power, hyperlocal climate events can elevate environmental consciousness significantly further. The air-quality in Beijing Tiananmen square is an example of a hyperlocal climate indicator (close to what was considered in the China section of this study). Indicators which capture hyperlocal information most relevant to each ETS community should be created (examples include California wildfires or Hurricanes and Nor-easters in the U.S. Eastern seaboard states).

Portfolio Approach
Once the allowance prices are being forecasted robustly, the fund must seek to put on exposures in each of the different allowances (either through spot or futures instruments). At this stage the following question arises - "How much of each ETS allowance asset should I buy?".

This is purely a portfolio selection or optimization problem and can be achieved using volume weights, liquidity weights, or weights chosen using other portfolio optimization techniques such as mean-variance, markowitz, or michaud optimization. Additional portfolio improvement techniques such as covariance matrix shrinkage (such as Ledoit Wolf, Rao Blackwellization, Oracle Approximating Shrinkage, Pedersen's Shrinkage method, Constant Residual Eigenvalue, Marchenko-Pastur) or clustering (Hierarchical Risk Parity, K-means clustering, K-medoids clustering, etc..) may be employed to improve return or risk metrics for the portfolio under actual backtesting.

Regardless of the approach chosen for portfolio construction, there is an important ***liquidity constraint*** to be additionally evaluated when actually forming the portfolio based on weights recommended by the different optimization methods. The ETF must account for the fact that it cannot take unlimited stakes in carbon allowances. The size of the carbon markets at present is only \$160 billion in total; With an average carbon price per ton of ~\$20, that amounts to a total of \$8 billion in tradable permits. The funds investment strategy if involves staking must hence assume that it can at max assume an upper-bound position of \$16 billion in carbon allowances, so as to not upset any local regulators and more importantly emitting companies. However, if the entire structure of the carbon markets change in the future (2030, 2035), it is possible to assume that the fund can take incrementally larger stakes upto 10% of the size of the carbon markets. This liquidity upper bound described above, turns the forecasting problem into a constraint optimization problem, where even if a portfolio optimization technique assigns a high weight for a particular carbon allotment index, a constraint equal to at-max 5% of the market size must be applied. Further, this upper-bound maximum percentage should be parameterized into the dynamic portfolio construction solution for rebalancing. Computationally, these constrained optimization problems can be represented as symbolic equations and solved using dynamic programming, non-linear optimization and symbolic computing packages in Python.

A final consideration, in choosing the portfolio composition would be around whether only stocking carbon allowance exposures shall be sufficient to attaining any environmental targets set out by the ETF (if any) with regard to banking of credits and thereby emissions reduction. This means that the portfolio manager would need to think of whether any other commodity underlyings may need to be commingled into the fund for sufficient attainment of decarbonization or energy transition goals of the fund.

Backtesting metrics
As described earlier a carbon ETF set out for itself not only a financial risk-adjusted return target, but also an environmental impact target (attained through the banking of allowances held by the

fund). The portfolio manager and quant will need to select the most robust backtesting metrics for the problem at hand. Some recommended metrics are as below -

- Return metrics: Percentage financial return
- Risk metrics: Sharpe Ratio, Maximum Drawdown
  The financial risk and returns shall be attained by programming buy-and-hold or buy-sell operations into the backtest based on the ETS price forecasts

- Environmental return metrics: CO_2e reduction
  The environmental returns shall be attained by programming in the sinking of ~10% of all allowances held by the fund every 3 or 5 years (returned back to the exchanges to remove from circulation). (The numbers 10%, and 3 and 5 years are to be parameterized by the backtest)

Commercialization and Market Perception

Even assuming that the entire quantitative and portfolio management layers of the proposed ETF is working well, i.e. there are no concerns across the data engineering, data validations, model development, model validation, quant-trading communication and backtesting layers, there are still risks associated in the process of commercialization and sale of the fund to investors. There may be skepticism associated with new proponents and a new product offering, and hence establishing the credibility in fund management and fund administration is as vital to the fund's success as are good forecasting methodologies.

5.3 Climate Scenarios

The motivation to discuss climate scenarios (also refer *Appendix G* for a summary of standard scenarios) in detail in this section, is to present the standard scientific consensus around how long-term climate forecasts and climate scenarios are developed, and how a new product like a carbon ETF seeking to explain its environmental impacts into the long term may lean on these estimates as opposed to devising other non-standard methods for scenario generation, not aligned to global goals.

Specifically, this subsection will discuss (i) '*Integrated Assessment Modeling*' which is a collection of several models which model sectoral emissions for different sectors at a country or regional level, and aggregate them to a global scale and (ii) '*Standard Scenarios*' defined by three leading authorities on climate projections namely - the Intergovernmental Panel on Climate Change (IPCC) based in Geneva, Switzerland, the International Energy Agency (IEA) based in Paris, and the Network for Greening the Financial System (NGFS), a Banking Supervision Committee for international banks, currently hosted by the Banque de France.

"**Integrated Assessment Models (IAMs)** are simplified representations of complex physical and social systems, focusing on the interaction between the economy, society and environment." This means that they could use economic and social variables to model environmental change, or the other two combinations. They come in three temporal forms - *perfect foresight* (implying the system has deterministic awareness of future states), *recursive dynamic* (myopic or make near-term decisions without foresight into future steps), and *imperfect adaptive* (where economic decisions are based on past, current and imperfectly anticipated future information). They also come in two economic forms - *general equilibrium* (where all economic agents and their interactions are fully represented in the model), or *partial equilibrium* (where only a subset of economic sectors or agents are represented). They include variables from the human-system, land-system, energy system, and climate system and generally classify as *cost-benefit* IAMs or *process-based* IAMs.

IAMs are also diverse with regard to their *sectoral* (energy, buildings, industry, maritime, power etc..) or *geographic coverage* (national, sub-national, regional, sub-regional). By virtue of their board definition, IAM literature also encompasses a variety of other known modeling techniques used by economists, such as macro-econometrics, macroeconomic-growth, life-cycle assessment, computable general equilibrium (CGE), factor productivity, structural change, technological spillover, trade flows, investment flows, strategic interaction and other models. Finally, IAMs come in two forms - top-down or bottom-up approaches. A *bottom-up approach* involves dividing the global climate system into a nested tree of sectors, sub-sectors and more specifically regionally, and evaluating detailed estimations of emissions based on more detailed sector specific assessments. E.g. in the building sector). Consequently, the bottom-up approach presents more systematic individual technological details about a reduced number of mitigation strategies of a specific sector or sub-sector; However, these models disregard relations between specific sectors/technologies and miss evaluating interactions with the whole system. A *top-down* approach is a more aggregated and global analysis, in detriment of less detailed technological heterogeneity. It tends to focus on interactions within the whole system, such as market and policy instrument interactions within the global economy systems. Due to the modeled interactions, top-down approaches tend to outperform bottom-up approaches and are better at predicting economic structural or regime changes.

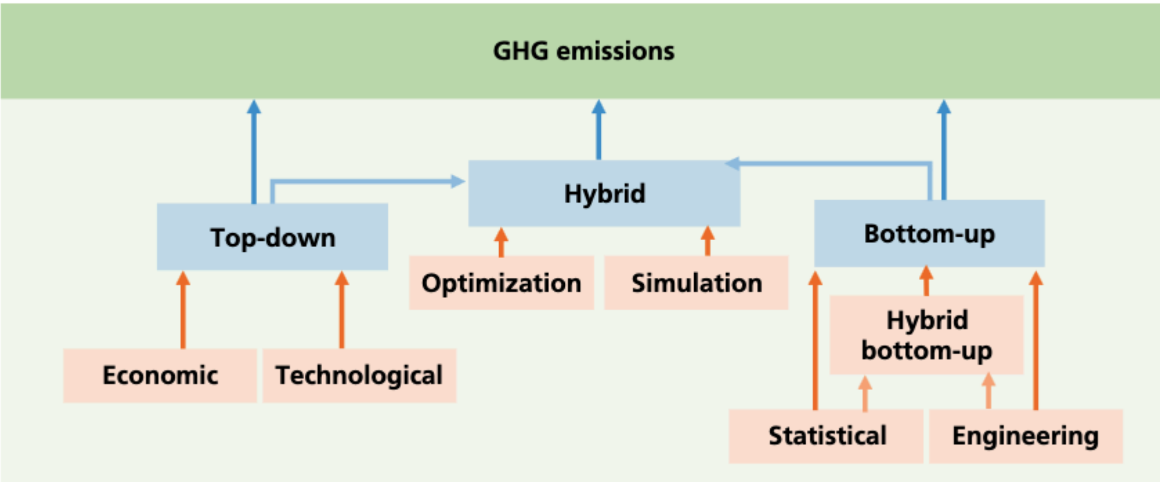


Figure 23 - Illustration of Top-down, Hybrid or Bottom-up modeling approaches
Source: *IPCC (Page 1849)*

With regards to using IAMs to model the entire global climate system and derive emissions trajectories into the long-term, just a single sector, such as energy, could have an entire suite of models which can be validated and aggregated to derive total sectoral emissions. The types of models for just the energy sector could be - (i) Modelling Electricity System Operation and Planning with Large-scale Penetration of Renewables (and its multiple regional variants) (ii) Modeling the interaction of different energy systems (including electricity, natural gas, hydrogen, hydro, thermal, gas etc..) to predict the evolution of the energy system into the long-term (iii) Hybrid models of energy and the macroeconomy forecasting energy demand, energy supply or energy transformation and their regional variations. Similarly there could be a multitude of models, modeling small interactions for the building sector, transport sector, industrial sector or the AFOLU (agriculture, forestry and other land-use sector. The combined emissions level across all sectors at a planetary scale is hence measured from submissions for smaller sectoral or regional models into a large scenario database managed by a technical scientific body such as the IPCC, IEA or NGFS.

While it would not be possible to delve into each IAM used to generate a forecast into the scenario database (typically there are hundreds of submissions), the next part of the literature review will be dedicated to sharing the specifics and interpretation of standard climate scenarios of the IPCC, IEA and NGFS which are used by different institutions to assess long-term impacts of emissions trajectory predictions from climate scenarios. (For a condensed tabular format refer *Appendix G*).

'Scenarios' are defined as descriptions of alternative future developments. They are used to explore how alternative courses of action could impact future developments. Climate change scenarios, specifically, are used to study future developments, to understand pathways leading to attainment of future climate goals, to look at integrated views of cross-domain research, and to inform society about individual choice and collective action. The rationale for having standard scenarios is to be able to compare the research between different groups.

The IPCC Scenarios (derived from a set of models called the CMIP models) explore two major long-term ideas - (i) How the GHG concentration in the atmosphere may change in relation to different radiative forcing (quantification of shifts in the Earth's atmospheric energy balance through anthropogenic emissions), through a set of standard scenarios called the '*Representative Concentration Pathways*' and (ii) Future greenhouse gas concentrations (2000-2100) in ppm under different projections of socioeconomic global change and 'qualitative' climate policies based on underlying factors such as population, technology, economic growth and education levels, through a set of standard climate scenarios called the '*Shared Socioeconomic Pathways*'. Each scenario begins in 2000, and ends with a different level of CO2 concentration by 2100 (end-century). Each of these different CO2 concentration levels, correspond to a certain temperature-equivalent (temperature rise from (from global warming) by 2100: <4 ℃).

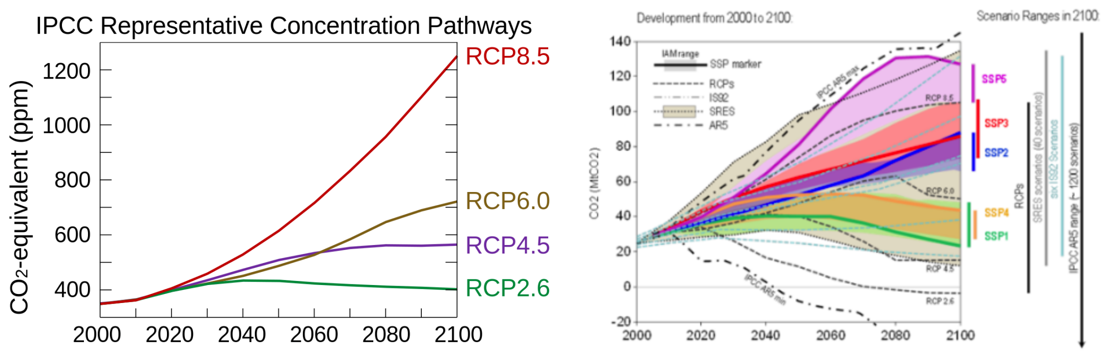


Figure 24 - IPCC Scenarios and associated end-century CO_2 concentration levels (Source: *SkepticalScience and UNFCCC***)**

IEA Scenarios

THE IEA Scenarios (derived from a set of models called the GEC models) explore how the global climate may evolve in relation to the evolution of the energy system over time. Its four scenarios titled NZE (Net Zero emissions by 2050), APS (Announced Pledges Scenario), STEPS (Stated Policies Scenario) and SDS (Sustainable Development Scenario) assume different levels of attainment of targets (defined as a combination of 2030 Paris Agreement, 2030 SDG, 2030 national announced of legislative targets, and 2050 long-term targets) by countries and their associated impact on GHG concentrations (GtCO2e) and temperature increases from global warming by end-century. NZE assumes full attainment of all targets - announced and future, APS assumes full attainment of self-determined targets but not externally enforced targets, STEPS assumes limited attainment based on practical realities of technological transfer and growth and evolution of the energy sector, and SDS assumes full attainment of the UN SDGs indicators and targets by 2030.

Most importantly, the gap between the NZE and APS scenarios (demonstrated below: gap between green and orange lines), is studied by climate researchers as the ‘Paris Agreement ambition gap’.

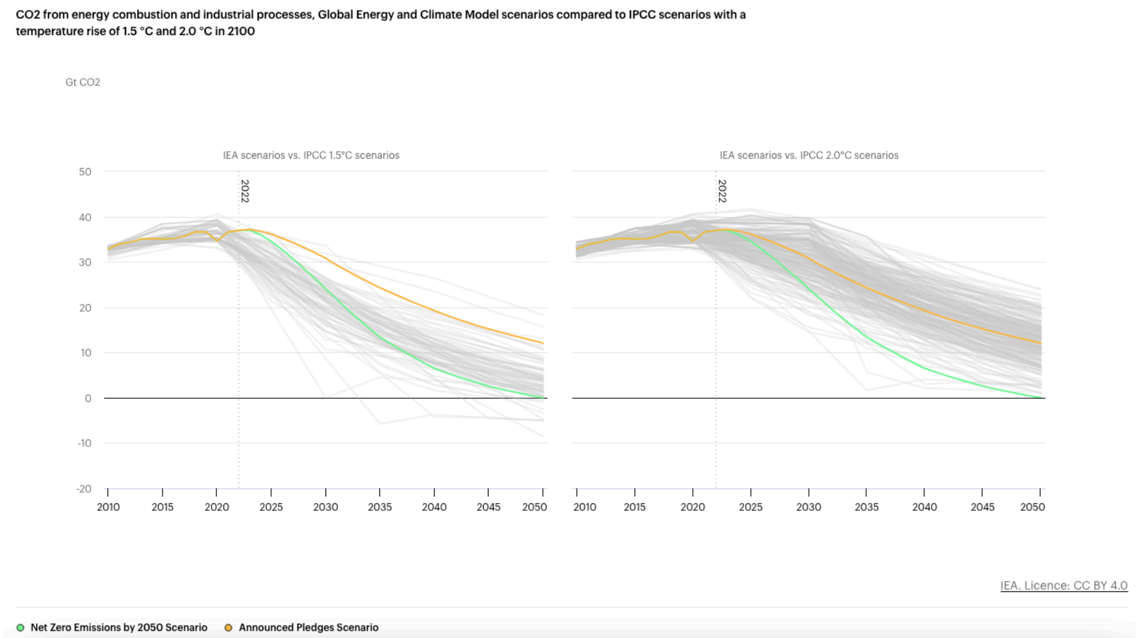


Figure 25 - IEA NZE and APS Scenarios - the Paris Agreement ambition gap (Souce: *IEA*)

NGFS Scenarios

The NGFS Scenarios (derived from a series of IAM models called REMIND MAgPIE, MESSAGE GLOBIOM, and GCAM) are more aligned to how the banking sector studies climate risk in two categories - climate change (physical risk) and climate policy and technology trends (transition risk), and study how the attainment of these targets, given the challenges to their attainment, could evolve differently into the future (upto 2050). Scenarios show a combination of higher and lower risk outcomes. Given the NGFS Scenarios are evaluated using IAM models, the models produce as outcomes the net GHG concentration, and economic variable forecasts (such as global GDP) into the long-term for adoption by banks for modeling into their climate risk measurements. The models are sophisticated and consider the implications of geopolitical risk, carbon prices and improved modeling of physical risks and hazards (such as heatwaves, droughts, river floods and tropical cyclones) at the downscaled country-level. Given that the goal of the NGFS scenarios is to help central banks and supervisors explore the possible impacts on the economy and the financial system, the NGFS is also developing a set of near-term scenarios 3-5 years, to help banks tangibly manage the financial impact of climate risks.

NGFS scenarios show that attainment of climate targets require ambitious transition efforts across all sectors of the economy and that immediate coordinated transition will be less costly; In the short run, mitigation costs will lead to the strongest negative impacts on GDP, with economic cost diverging significantly after 2040.

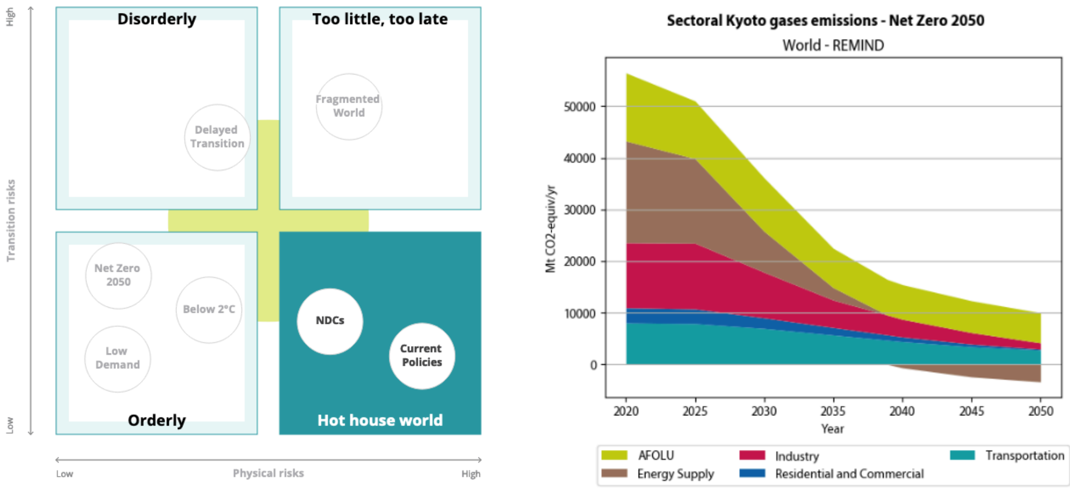


Figure 26- NGFS Scenarios and breakdown of emissions by sector up to 2050 from REMIND model
(Source: *NGFS.net and NGFS Climate Scenarios Technical Documentation***)**

5.4 Long Term Forecasts

Investors into this fund are expected to be corporations, governments, funds and other entities, aligned with global climatic targets, and seeking to achieve a risk-adjusted financial return and an environmental impact through their investment. Any such institutional investor hence is expected to ask of the fund manager of a carbon ETF -

"As an institution, I invested $X in this product, how would I be contributing to broader climatic goals of 2030 or 2050?"

In seeking to answer this question, it is imperative that the fund not only carry out backtesting on short time-scales (such as the 2017-2023 period in our study), but also be able to discuss market evolution, macroeconomic evolution and climate scenarios (evolution of emissions) into the long term. Specifically, they must also be able to report their own alignment with international climate targets such as '(Paris Agreement NDC targets) 2030' or '(Net zero by) 2050' etc. Since it is not possible to validate such long-term forecasts, this can be viewed as a quantitative or stochastic **simulation** problem.

In order to construct long-term forecasts, our datasets would need to be projected forward out to 2030 or 2050. Each covariate would need to be project forward using either of the following approaches -

1. Expert judgment and priors
2. Statistical forecasting of a single vector - using ARIMA, stochastic modeling or modal decomposition
3. Statistical forecasting of multiple vectors - using correlation matrices of a collection of vectors and performing cholesky decomposition as with the stochastic forecasting of multiple related market instruments
4. Outputs from long-term macroeconomic or climate models and standard scenarios
5. Awareness and anticipation of market structure changes, such as future planned sectoral inclusion, sectoral reduction through integration into national ETS or other treaty (such as CORSIA), or increase in covered entities and emissions

Once the above is established and a long-term dataset has been constructed, the best identified model from model training can be then trained using a rolling window of training / test data to establish predictions of financial and environmental return out into the long-term. Intuitively and as guided by the climate scientific bodies, we can anticipate that it is technological change that can cause the greatest shift in the emissions level in the long-run.

6. CONCLUSION

This paper has demonstrated an approach to forecast spot carbon allowance prices across multiple exchanges, and using multiple models simultaneously. The paper has also discussed several methods to improve model selection and performance, and several challenges that are to be addressed before a set of robust forecasts can actually power a commercial carbon ETF.

APPENDIX A: LITERATURE REVIEW - EMISSIONS TRADING SYSTEMS (Source: ICAP)

Emissions Trading System (ETS)	**RGGI**	**California**	**Korea ETS**
Founded	2009	2012	2015
Governed by	RGGI Inc. (501c3)	California Air Resources Board	Ministry of Environment Ministry of Economy and Finance Korea Exchange (KRX) Greenhouse Gas Inventory and Research Center (GIR)
Covers	Connecticut, Delaware, Maine, Maryland, Massachusetts, New Hampshire, New Jersey, New York, ~~Pennsylvania~~, Rhode Island, Vermont and ~~Virginia~~	California	Republic of Korea
Goals	30% reduction from 2020 levels by 2030	40% GHG emission reduction from 1990 levels by 2030	To establish a market-based mechanism to meet the Paris Agreement 2030 NDC target (reducing GHG emissions by 40% from the 2018 level by 2030), and the country's objective to become carbon neutral by 2050
GHGs covered	CO_2	CO_2, CH_4, N_2O, SF_6, HFCs, PFCs, NF_3, and other fluorinated GHGs	CO_2, CH_4, N_2O, HFCs, PFCs, SF_6
Sectoral Coverage	Power	Transport, Buildings, Industry, Power	Maritime,Waste, Domestic Aviation, Transport, Building, Industry, Power
Phases	Ph1 (2009-2011) Ph2 (2012-2014) Ph3 (2015-2017) Ph4 (2018-2020) Ph5 (2021-2023) **Ph6 (2024-2026)**	First (2013-2014) Second (2015-2017) Third (2018-2020) Fourth (2021-2023) **Fifth (2024-2026)**	Ph1: 2015 to 2017 Ph2: 2018 to 2020 **Ph3: 2021 to 2025**
Cap	63 $MtCO_2e$	280.6 $MtCO_2e$	567.1 $MtCO_2e$ (2024)
Covered Entities	228	~400	804 (2023)
Allowance Allocation	Auctioning	Free Allocation + Auction	Free Allocation + Auction
Average auction price	USD 12.81	USD 41.76 (Feb 2024)	USD 8.17 (2022)
Auctioned share (%)	~92%	~70%	3%
Total Revenue since inception	USD 7,160 million	USD 26.97 billion	USD 845.2 million
Use of Revenues	energy efficiency, direct bill assistance, beneficial electrification, GHG abatement, and clean and renewable energy	35% must benefit disadvantaged and low-income communities; Otherwise, investments into environmental, economic, and public health projects across the state.	emissions mitigation infrastructure, low-carbon innovation, and technology development for small-and mid-sized companies
Use of offsets	3.3% of an entity's liability	up to 8% of an entity's obligations	KOCs and international credits converted to KCU are eligible
Linkages	All RGGI States	Quebec and Ontario (WCI)	None
Compliance mechanism	1 allowance per $1tCO_2e$	1 allowance per $1tCO_2e$	1 KAU represents 1tCO2e)
Compliance period	Yr. 1 - 50% of annual emissions, Yr. 2 - 50% of annual emissions, Yr. 3 - remainder	Yr. 1 - 30% of annual emissions, Yr. 2 - 30% of annual emissions, Yr. 3 - remainder	Entities must surrender allowances for the previous emissions year by the end of August
Penalty for non-compliance	3x the number of allowances	4x the number of allowances	Min (3x average market price of allowances of the given compliance year, USD 76.58 per tonne)
Market Stability Provisions	Auction price floor, Cost containment reserve (CCR), Emissions containment reserve (ECR)	Auction Reserve Price, Allowance Price Containment Reserve (APCR)	Triggers (market intervention for price control), Instruments (ad-hoc auctions, limits on allowances / borrowing / offsets)

Emissions Trading System (ETS)	EU-ETS	Beijing Pilot ETS	Shanghai Pilot ETS
Founded	2005	2013	2013
Governed by	European Commission	Beijing Municipal Ecology and Environment Bureau (EEB) (北京市生态环境局)	Shanghai Municipal Ecology and Environment Bureau (EEB) (上海市生態環境局)
Covers	Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden	Beijing metropolitan area	Shanghai metropolitan area
Goals	reducing GHGs by ~55% from 1990 levels by 2030, and attaining climate neutrality by 2050	reducing CO_2 emissions by 10% or more compared to 2030 modeled peak emissions, and by ~18% compared to 2020 measured emissions levels	reducing CO_2 emissions by 5% compared to peak (2025) levels by 2035 and attaining full climate neutrality by 2060
GHGs covered	CO_2, HFCs, N_2O, PFCs, SF_6	CO_2 only	CO_2 only
Sectoral Coverage	Maritime, Domestic Aviation, Industry, Power	Transport, Buildings, Industry, Power	Maritime, Domestic Aviation, Transport, Buildings, Industry, Power
Phases	Ph1: 2005 to 2007 Ph2: 2008 to 2012 Ph3: 2013 to 2020 **Ph4: 2021 to 2030**	**2013 and ongoing**	Ph1: 2013 to 2015 (trial phase) **Ph2: 2016 to present**
Cap	1,442 $MtCO_2e$ (2024)	44 MtCO2e (2022)	100 MtCO2e (2022)
Covered Entities	8,640 energy and industry installations and 309 aviation and maritime transport companies (2022)	909 entities (full-obligations), 388 (reporting only)	357 (2022)
Allowance Allocation	Free Allocation + Auctioning	Grandparenting (based on historical emissions or emissions intensity in the past three years), Benchmarking (specific allowance capacities for new entrants and sectors), Auctioning	Grandparenting (based on historical emissions or emissions intensity in the past three years), Benchmarking (specific allowance capacities for new entrants and sectors), Auctioning
Average auction price	USD 90.25 (2023)	USD 16.26 (2023)	USD 8-10
Auctioned share (%)	~57%	~5%	3%
Total Revenue since inception	USD 206 billion	USD 38.67 million	USD 64.41 million
Use of Revenues	goes into the budgets of member states, 50% of which must be mandatorily used as state aid to climate and energy initiatives and reported to the European Commission	go into the city treasury from where it is used for municipal projects	go into the provincial treasury from where it is used for municipal projects
Use of offsets	not allowed since phase 4	CCERs (2013 onwards) can be used upto a maximum of 5%	CCERs and SHCERs can be used upto a maximum of 5%
Linkages	Swiss ETS (since 2020)	None	None
Compliance mechanism	1 allowance per $1tCO_2e$	1 allowance per $1tCO_2e$	1 allowance per $1tCO_2e$
Compliance period	expected to submit emissions monitoring plans, reports and have them verified by independent verifiers by the end of the year (March)	surrender allowances by mid-June of the following year	surrender allowances by mid-June of the following year
Penalty for non-compliance	EUR 100 (USD 108.13), adjusted for inflation, for each tCO_2 emitted for which no allowance has been surrendered	fines of up to CNY 50,000 (USD 7,058) + 5x average market price over the previous six months for each missing allowance	USD 1,411-USD 7,058 for failure to submit, USD 7,058-USD 14,116 for non-compliance
Market Stability Provisions	Backloading or withholding allowances, Market Stability Reserve (MSR) to influence price stability or emissions control	A reserve mechanism managed for price stability, and limits on price volatility (+-20%) beyond which trading may be suspended and position size limited.	A reserve mechanism managed for price stability, and limits on price volatility (+-30%) beyond which trading may be suspended and position size limited.

APPENDIX B: LITERATURE REVIEW - CARBON PRICE FORECASTING

Statistics-based forecasting		
Feature Space (Y, Xi)	**Modeling Paradigms**	**Results / Model Performance**
Benz and Truck (2008) focusing on EU-ETS		
Y: Log returns from a composite price signal constructed from spot, forwards and futures as independent variables; Xi: No features or feature engineering otherwise.	Markov transition and AR-GARCH based Training period: 2005, Testing period: 2006	Adequate CO2 process does not necessarily follow seasonal patterns, and martingale property, and time and price dependent volatility structure is assumed. Traders are interested in modeling long-term spot prices, but also short-term / intraday futures prices.
Chevallier (2008) focusing on EU-ETS		
Y: Returns of ECX carbon futures; Xi: Equity Dividend yields, Junk bond premium, UST yields, Excess return of CRB Index over UST, Power demand, Energy prices	ARMAX-ARCH models Data: Phase II (2008-2012) of EU-ETS (because Carbon futures returns exhibit non-zero skewness and excess kurtosis (fat tailed leptokurtic))	EU-ETS is operating as a very specific commodity market Dividend yields, Junk bond premium, energy prices and market excess returns have high explanatory power; Carbon price is not highly connected linked to macro variables because power producers have fuel switching behavior
Byun and Cho (2013) focusing on EU-ETS		
Y: EU-ETS EUA Carbon futures prices from ECX; Xi: Carbon option implied volatility, volatility of Brent, coal, natural gas and electricity	GARCH, EGARCH, TGARCH and GJR-GARCH with normal and t-distributions with AR, MA and GARCH terms varying from one to three	Supplementing the feature set with energy volatilities improves predictions; Carbon options are not a good predictor due to low trading volume. Based on loss functions such as MSE, MSE-LOG, MAE, MAE-LOG, and QLIKE complex GARCH specifications are rejected, and residuals approach a normal distribution.
Zhu and Wei (2013) focusing on EU-ETS		
Y: EU ETS Futures; Xi: None (purely statistical model working with the carbon price)	Least squares SVM or ANN with ARIMA with Gray Model Correlation, with particle swarm optimization for optimal LSSVM kernel parameters	The hybrid ARIMA-LSSVM and ARIMA-ANN models capture linear and non-linear components of the carbon price adequately. RBF kernel function performs the best for LSSVM.
Yang et al (2021) focusing on EU-ETS, and China ETS		
Y: EU ETS Carbon Allowance Spot price; Xi: None (purely statistical model working on decomposition of the carbon price signal)	ICEEMDAN (Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) data pre-processing and GBiLSTM, CNN and ELM (extreme learning machine) as second stage ML predictor; Multiobjective dragonfly optimization algorithm (MODA) is used as weighting algorithm to ensemble sub-models.	Best MAE ~0.0005 (EU) 0.013 (Beijing) 0.0448 (Shenzen)
Machine / Deep Learning-based forecasting		
Feature Space (Y, Xi)	**Modeling Paradigms**	**Results / Model Performance**
Fan et al (2015) focusing on EU-ETS		
Y: EU-ETS EUA Carbon futures prices from ECX Xi: Extreme weather, macroeconomic variables, EUR/USD exchange rate	Identifying chaotic price elements using the maximum Lyapunov exponent, the fractal correlation dimension and Kolmogorov entropy and reconstructing signal; MLP with k-fold CV as second-stage model. Data: Phase III (2013-2015) of EU-ETS.	Carbon prices are found to be chaotic rather than stochastic; The MLP model possesses good performance in both level and directional measurement.
Atsalakis (2016) EU-ETS		
Y: EU-ETS EUA spot price Xi: none	3 Computational Intelligence models: PATSOS: Neuro-fuzzy controller with closed loop, an ANN, an adaptive neuro-fuzzy inference system (ANFIS)	RMSE, MSE and MAE were used as loss functions. Simulated trading results based on prediction vs. buy and hold. PATSOS outperformed other models.
Zhang et al (2017) focusing on computational methods		
Y: Electric load (New South Wales, Australia) Xi: N/A (Predominantly a paper to create a more robust computational forecasting method) Data: Apr - Jun 2009	(CS (cuckoo search) - SSA (single spectrum analysis) - SVM (support vector machine) for electric load forecasting used in studying the operation of power systems. CS is a replacement for grid search to search for SVM optimal parameters. SSA, is a technique to extract trends from time series.	MAPE, MSE and MAE were used as loss functions. CS-SSA-SVM outperforms SVM, CS-SVM, SSA-SVM, SARIMA and BPNN.
Lu and Ma (2020) focusing on Beijing, Shenzhen, Guangdong, Hubei, Shanghai, Tianjin, Chongqing (2013 onwards) pilots, Fujian (2016 onwards)		
Y1: Carbon Allotment Prices Y2: Carbon Allotment Trading Volumes Xi: Price, Volume, Policy decisions or Regulatory market adjustment	CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) data denoising with XGBoost, Random Forest, kernel-based nonlinear extension of	Average accuracy of CEEMDAN-RBFNN and CEMDAN-GWO-KNEA in carbon price prediction is

(categorical)	the Arps decline model optimized by grey wolf optimizer (GWO-KNEA), SVMs optimized by particle swarm optimizer, and fruit fly optimizer and simulated annealing algorithm, and radial basis function neural network (RBFNN) based predictions.	98.40% and 97.89%, and they are the best performing models.
Zhang and Wen (2022) focusing on EU-ETS		
Y: EUA Futures Closing Price Xi: Dow Jones Euro Stoxx 50 Index, crude oil price, natural gas price (Henry Hub's natural gas future price), and coal price (Australian BJ thermal and Australian Newcastle), Training: 2005-04-22 to 2013-02-13 Test: 2013-02-14 to 2017-12-29	TCN-Seq2Seq deep learning model which focuses on the "sequence to sequence" layout to learn temporal data dependencies using only fully convolutional layers and predictive patterns, which are common in domains of carbon price forecasting.	Deep learning models on average tend to perform shallow machine learning models. The model outperforms all other benchmark models (ARIMA, RF, Xgboost, SVR, LSTM) with the highest DA (directional accuracy) value (0.9697), the lowest MAPE value (0.0027), and the lowest RMSE value (0.0149). Not necessarily does machine learning always outperform time-series models.
Rudnik et al (2022) focusing on EU-ETS		
Y: EUA day ahead spot price Xi: A range of 37 fuel and energy indicators traded on the ICE (analyzed in groups of 7 - power, natural gas, coal, crude, heating oil, unleaded gasoline, gasoil), and reduced to 4 vectors (explaining 99.6%) using PCA (using SVD algorithm)	Regression trees, Ensembles of Regression trees, Gaussian Process Regression (GPR) models, Support Vector Machines (SVM) models and Neural Network Regression (NNR) models.	Non-isotropic exponential Gaussian Process Regression model turned out to be the most advantageous and its price prediction can be considered very accurate (Best OoS MAE: 1.0618). Additionally, weather (extreme temperature and radical weather phenomena), industrial production levels and clean energy indices are known to impact the carbon price.
Yun et al (2022) focusing on EU-ETS		
Y: EUA Futures Prices (chosen due to larger volume and more liquidity) Xi: None (purely signal decomposition) Data: Jun 2, 2009 - Nov 23, 2021	CEEMDAN (Complete Ensemble Empirical Mode Decomposition with Adaptive Noise) data denoising with LSTM Forecasting model.	Established the general consensus that GARCH cluster > implied volatility > Knn based models; Concluded that CEEMDAN is preferred for carbon price due to the benefits of signal mode decomposition at different time-scales. RMSE 0.638342, MAE 0.448695, MAPE 0.015666 and DA 0.687631 which beats benchmarks, and short term forecasts are more accurate.
Wang, Jiang and Shu (2022) focusing on China National ETS		
Y: Closing allowance price data from CEEX Xi: None (though they concur with energy prices, and macro-control variables as explanatory factors for China) Data: Dec 19, 2013 to Aug 11, 2022	SSA (singular spectrum analysis) + LSTM	Concurs that hybrid and mixed model approaches are the best for carbon price forecasting due to multi-frequency, non-linearity and non-stationarity. RMSE and MAE are used as the loss functions. The model outperforms other benchmark models (ARMA, Random Forest) with RMSE 3.9 and MAE 0,0456
Shahzad et al (2023) focusing on ICE Futures - underlyings are EU-ETS, WCI, RGGI and UK ETS allowances		
Y: ICE Carbon Emissions Futures Xi: WTI NYMEX Futures, Natural Gas Futures and (Dow Jones) Heating Oil Futures Data: 2005 to 2018	SVM, MLP, LSTM vs. a multiple regression benchmark	Carbon prices are correlated with social vulnerability and energy poverty. LSTM is best performer with MAPE 31% and MSE of $0.4
Wang et al (2023) focusing on Hubei ES		
Y: Hubei Emission Allowances (HBEA) spot prices Xi: 23 features representing the impact of the environment, energy and industry, economy and carbon market	CEEMDAN for modal decomposition PCA for dimensionality reduction Dual-stage attention RNN (DARNN) encoder-decoder called Seq2Seq.	Commodity prices, financial market tickers, macroeconomic variables, extreme weather, economic sentiment and economic development indices affect the carbon price. One significant flaw in a basic LSTM model is its inability to capture the connection between various input characteristics and its focus solely on capturing the temporal fluctuation of the input time series data unlike Seq2Seq. The Model (MAE = 0.75, MAPE = 1.59, RMSE = 1.28) outperforms benchmarks.
Feng et al (2023) focusing on Guangdong, Hubei, and Fujian		
Y: Emissions allowances prices (from Choice financial terminal, and Wind database) Xi: Brent Crude Oil CFD, Bohai Sea Power Coal Price Index, Natural Gas Market Quotation (min-max scaled), Shanghai Stock Exchange Industrial Index, DJIA, S&P500, CSI300, EU-ETS EUA, Max Temp, Min Temp, Air Quality Index (Data: 2013-11 to 2020-04)	PACF and Random Forest are used initially for feature selection based on Gini importance GWO (grey-wolf optimizer)–XGBOOST–CEEMDAN CEEMDAN for modal decomposition XGBoost for prediction GWO for XGBoost parameter optimization	Carbon price research can be categorized as models based on historical data and models based on influencing factors. The Hybrid model outperforms benchmarks based on loss functions such as MSE, MAE, RMSPE, MAPE, R^2 Guangdong MAE 0.22, Hubei MAE 0.05, Fujian MAE 0.05.
Shi et al (2023) focusing on Shenzhen		

Y: Shenzhen allowance prices Xi: n/a	CNN-LSTM with batch size 32, 1 CNN layer, time step 10, 50 neuron LSTM, and 100 and 150 epochs in training.	CNN-LSTM model achieved a level of accuracy comparable to other popular models such as CEEMDAN, Boosting, and GRU, 40% training speed improvement, 9% reduction in MSE and a 0.0133 shortening of the Z-score range. (Z-score can assess the model's robustness in coping with anomalies, outliers, and departures from the predicted pattern in the time series by measuring the presence and characteristics of outliers)
Zhang et al (2023) focusing on EU-ETS, and Beijing, Shanghai, Guangdong, Shenzhen and Hubei pilots.		
Y: Allowance prices (EUA, BEA, SHEA, GDEA, SZA and HBEA) Xi: COVID-19, Macro-variables (GDP, WTI, CSI300, Coal price), number of new confirmed COVID-19 cases per day, new environmental goal or treaty Data: May 2014 to Jan 2022	SSA for trend and period decomposition followed by six ML/DL predictions - SVR, LSSVR, Random Forest, XGBoost, RNN and LSTM	Other researchers have also used a 'conflict indicator' capturing the effect of modern conflicts (such as Russia-Ukraine) on the Carbon Allowance Price. The model is able to capture various effects of the COVID-19 pandemic; SSA-SVR and SSA-LSSVR perform the best in the context of EU-ETS and China based on RMSE and MAE. EUA and Hubei prediction have the lowest error amongst the six markets.
Dong et al (2023) focusing on Hubei, Guangdong, and Shanghai		
Y: HBEA, GDEA and SHEA allowance prices from China Emissions Exchange Xi: n/a (only decomposition of signal based forecast) Training: Apr 2, 204 to Nov 4, 2022 Test: Nov 7, 2022 to Dec 29, 2022	Lp-CNN-LSTM model: Carbon price forecasted using to strategies namely CEEMDAN-CNN-LSTM and Multi-CNN-LSTM (using technical indicators construed from OHLC prices, and Volume information) and ensemble using the Lp norm	MSE, RMSE, MAE and MAPE were used as the loss functions. MAE for CEEMDAN-CNN-LSTM was 0.97, and for Multi-CNN-LSTM was 0.85 and for the ensembled Lp-CNN-LSTM was 0.78 for Hubei. Predictions were better for Hubei, as compared to Guangdong, as compared to Shanghai (in that order).
Peng et al (2023) focussing on Shenzhen		
Y: SZA (log-diff, and first diff) (has large scale and more complete data since first carbon market) Xi: ~53 explanatory features including - Commodity variables (energy, non-ferrous metals, agri products), Stock and Bond Market Variables, Economic and Industry Composite Variables	Three ML paradigms SVR, Random Forest and XGBoost, ensembled with a double-sliding-window method, with walk forward validations and 7 base learners Train: March 27, 2018 to May 9, 2021 Test: May 10, 2021 to July 6, 2022	Linking carbon market vol with carbon return forecasting improves predictions. MSE, MAE and Huber loss were used as loss functions. Discussed Investment portfolio construction (Section 2.3.4), which measures portfolio return using Sharpe Ratio and a utility function to holding the carbon portfolio over 1y treasury called the Certainty Equivalent Return. Homogeneous or heterogeneous ensembles with SVR as meta model performed the best.
Recessary EU (2023) focusing on EU-ETS		
Y: EUA XI: n/a	Black-box, unavailable	As of Jan 2024, EU Allowances (EUAs) are forecasted to average at 74.11 EUR/MT in 2024 and 83.31 EUR in 2025, marking a decrease of 11.3% and 6.3%, respectively, from the forecasts made in October.
Schjoiset, S (2023) focusing on EU-ETS		
Y: EUA spot price Xi: Market Balances (MT) (supply and demand of credits)	Expected to be a Dynamic Stochastic General Equilibrium Model (but no further detail available)	The forecast studied was a 2021 forecast made in 2014; The model forecast 40% GHG reduction, and 2.2% cap reduction factor for 2021.

APPENDIX C: DATA DICTIONARY

H - Human factors, I - Institutional factors, M - Market factors, T - Technological factors, E - External factors, S - Static data (categorical)

1. **Regional Greenhouse Gas Initiative (RGGI) allowance price prediction - Feature Set**

	Factor type	Variable	Units	Frequency	Source(s)	Data Management
Target (Y)	n/a	Allowance Price	USD / ton	daily	Bloomberg: RGGISPOT index supplemented with prices from ICAP Allowance price explorer	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
Feature (Xi)	S	Compliance Period Identifier	n/a	various	RGGI Inc.	'Third', 'Fourth' or 'Fifth' based on dates. This is also encoded as its own categorical variable.
	I	Market Stability Action	tCO2e	ad-hoc	RGGI Inc. and ICAP Factsheet	Indicator to show if reserve was utilized to regulate price or emissions. Denoted as amount of allowances bought or sold
	I	Number of new free allowances Introduced	#	annual	RGGI Inc. Market Monitor Reports	Forward fill, followed by backfill to make into continuous series
	I	Number of new auctioned allowances	#	monthly	RGGI Inc. Market Monitor Reports	Forward filled between auction dates
	I	Weighted average sectoral abatement difficulty	Index (0,1)	ad-hoc	RGGI Inc. Market Monitor Reports and ICAP Factsheet	An index between 0 and 1 prepared by weighting sectoral additions using a sector weights dictionary; Only power sector in the context of RGGI
	I	Traded Volume	#	quarterly	RGGI Inc. Market Monitor Reports (Figure 4)	Pandas linear interpolation for the unavailable dates (includes physical delivery and futures contract)
	I	Cap	tCO2e	annual	RGGI Inc. Elements of RGGI	Forward filled; Cap is set in advance, but carried for the entire year
	I	Adjusted Cap	tCO2e	annual	RGGI Inc. Elements of RGGI	Forward filled; Cap is set in advance, but carried for the entire year
	I	Annual Cap Reduction	tCO2e	annual	RGGI Inc. Elements of RGGI	Cap (previous year) - Cap (current year)
	I	Penalty for Non-Compliance	USD / ton	ad-hoc	'Compliance Enforcement' on ICAP Factsheet	3 * allowance price
	I	Regime Name	n/a	ad-hoc	National Governors Association (NGA)	One field per RGGI state showing the names of the Governors between the dates they served in the respective state
	I	Regime Type	Index (0,1)	daily	Computed using National Governors Association (NGA) data	Score assigned to each governor (regime name) based on their political party affiliation. 0 - Right Wing, 0.5 - Centrist and 1 - Left Wing. Overall regime type is calculated as the population weighted regime type index between 0 and 1

	M	Population	#	annual	USA Facts (from Census Bureau, until 2022) and macrotrends (2023)	One field per RGGI state Backfilled from annual reported to daily frequency
	M	GDP	USD	quarterly	USA Facts (from Census Bureau) and bea (2023)	One field per RGGI state Backfilled from quarterly reported to daily frequency
	M	RGGI Population	#	annual	USA Facts (from Census Bureau, until 2022) and macrotrends (2023)	Computed as the sum of individual populations of RGGI states
	M	RGGI GDP	USD	quarterly	USA Facts (from Census Bureau) and bea (2023)	Computed as the sum of individual GDPs of RGGI states, backfilled from quarterly reported to daily frequency
	M	GDP/Capita	USD	quarterly	USA Facts (from Census Bureau) and macrotrends (2023) and bea (2023)	Computed from quarterly nominal GDP, and prevailing population estimate for the quarter
	M	Inflation	%	quarterly	USA Facts (from Census Bureau) and bea (2023)	Computed from GDP and Real GDP through GDP Deflator
	M	Recession Indicator	binary	ad-hoc	n/a computed binary variable	1 if GDP decreases for three consecutive prior quarters, 0 otherwise
	I	Other Carbon tax	#	ad-hoc	RGGI Inc. (State Statutes)	Number of other prevailing carbon tax in the RGGI states during the period of the permit regulation
	E	Number of extreme climate events during the period	#	ad-hoc	NOAA NWS	Backfilled from available values
	E	Average Global Temperature Anomaly in Celsius	degC	monthly	NOAA	Backfilled from available values
	T	Cumulative VCM market surplus by sector (Issuance minus retirement) tCO2e	tCO2e	monthly	Quantum Commodity Intelligence	Backfilled from available values

2. **California ETS allowance price prediction - Feature Set**

	Factor type	Variable	Units	Frequency	Source(s)	Data Management
Target (Y)	n/a	Allowance price	USD / ton	daily	CARB allowance price image, from Argus.Inc. via automeris.io and ICAP Carbon supplemented	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
Feature (Xi)	S	Compliance Period Identifier	n/a	various	CARB Compliance Reports	‘Second’, ‘Third’ or ‘Fourth’ based on dates. This is also encoded as its own categorical variable.
	I	Market Stability Action	tCO2e	ad-hoc	CARB Program Data and ICAP Factsheet	Indicator to show if reserve was utilized to regulate price or emissions. Denoted as amount of allowances bought or sold
	I	Number of new free allowances Introduced	#	annual	CARB Allocated Allowances	Forward fill, followed by backfill to make into continuous series

	I	Number of new auctioned allowances	#	quarterly	CARB Auctioned Data	Forward filled between auction dates
	I	Weighted average sectoral abatement difficulty	Index (0,1)	ad-hoc	CARB Allowance Allocation and ICAP Factsheet	An index between 0 and 1 prepared by weighting sectoral additions using a sector weights dictionary
	I	Traded Volume	#	quarterly	CARB Summary of Market Transfers Report	Pandas linear interpolation for the unavailable dates
	I	Cap	tCO2e	annual	CARB Allowance Allocation	Forward filled; Cap is set in advance, but carried for the entire year
	I	Annual Cap Reduction	tCO2e	annual	CARB Allowance Allocation	Cap (previous year) - Cap (current year)
	I	Penalty for Non-Compliance	USD / ton	ad-hoc	'Compliance Enforcement' on ICAP Factsheet	(3 + 1) * allowance price
	I	Regime Name	n/a	ad-hoc	National Governors Association (NGA)	Names of the Governors between the dates they served in the state
	I	Regime Type	Index (0,1)	daily	Computed using National Governors Association (NGA) data	Score assigned to each governor (regime name) based on their political party affiliation. 0 - Right Wing, 0.5 - Centrist and 1 - Left Wing.
	M	Population	#	annual	USA Facts (from Census Bureau, until 2022) and macrotrends (2023)	Backfilled from annual reported to daily frequency
	M	GDP	USD	quarterly	USA Facts (from Census Bureau) and bea (2023)	Backfilled from quarterly reported to daily frequency
	M	GDP/Capita	USD	quarterly	USA Facts (from Census Bureau, until 2022) and macrotrends (2023) and bea (2023)	Computed from quarterly nominal GDP, and prevailing population estimate for the quarter
	M	Inflation	%	quarterly	USA Facts (from Census Bureau) and bea (2023)	Computed from GDP and Real GDP through GDP Deflator
	M	Recession Indicator	binary	ad-hoc	n/a computed binary variable	1 if GDP decreases for three consecutive prior quarters, 0 otherwise
	I	Other Carbon tax	binary	ad-hoc	CARB FAQ	If there was any other prevailing carbon tax in California during the period of the permit regulation
	E	Number of extreme climate events during the period	#	ad-hoc	NOAA	Backfilled from available values
	E	Average Global Temperature Anomaly in Celsius	degC	monthly	NOAA	Backfilled from available values
	T	Cumulative VCM market surplus by sector (Issuance minus retirement) tCO2e	tCO2e	monthly	Quantum Commodity Intelligence	Backfilled from available values

3. Korea ETS (K-ETS) allowance price prediction - Feature Set

	Factor type	Variable	Units	Frequency	Source(s)	Data Management
Target (Y)	n/a	Allowance price	USD / ton	daily	ICAP Carbon Allowance Price Explorer	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
Feature (Xi)	S	Compliance Period Identifier	n/a	various	ICAP K-ETS Factsheet	'Second', 'Third' or 'Fourth' based on dates. This is also encoded as its own categorical variable.
	I	Market Stability Action	tCO_2e	ad-hoc	ICAP K-ETS Factsheet	Indicator to show if reserve was utilized to regulate price or emissions. Denoted as amount of allowances bought or sold
	I	Number of new free Allowances Introduced	#	annual	GIR Library Annual K-ETS Reports and Biennial Updates	Forward fill, followed by backfill to make into continuous series
	I	Number of new auctioned allowances	#	semi-annual	GIR Library Annual K-ETS Reports and Biennial Updates	Forward fill, followed by backfill to make into continuous series
	I	Weighted average sectoral abatement difficulty	Index (0,1)	ad-hoc	GIR Library Annual K-ETS Reports and Biennial Updates	An index between 0 and 1 prepared by weighting sectoral additions using a sector weights dictionary
	I	Traded Volume	tCO_2e	quarterly	GIR Library Annual K-ETS Reports and Biennial Updates	Quarterly until 2020; Appropportioned quarterly from annual for 2021 and 2022; Forecasted at 3% annual growth for 2023.
	I	Cap	tCO_2e	annual	GIR Library Annual K-ETS Reports and Biennial Updates	Forward filled; Cap is set in advance, but carried for the entire year
	I	Annual Cap Reduction	tCO_2e	annual	GIR Library Annual K-ETS Reports and Biennial Updates	Cap (previous year) - Cap (current year)
	I	Penalty for Non-Compliance	USD / ton	ad-hoc	ICAP K-ETS Factsheet	Max (KRW 100,000 (USD 76.58) per tonne, 3 * allowance price)
	I	Regime Name	n/a	ad-hoc	Wikipedia	Names of the Prime Ministers between the dates they served in the respective state
	I	Regime Type	Index (0,1)	daily	Computed using Regime name *(Independent = 0.5, Democratic = 0.75, People Power = 0, Liberty Korea = 0)*	Score assigned to each regime based on their political party affiliation. 0 - Right Wing, 0.5 - Centrist and 1 - Left Wing.
	M	Population	#	annual	World Bank and focus-economics.com	Backfilled from annual reported to daily frequency
	M	GDP	USD	quarterly	FRED (NGDPNSAXDCKRQ) and Retuers	Backfilled from quarterly reported to daily frequency
	M	GDP/Capita	USD	quarterly	World Bank Open Data and statista	The values from WBOpen Dat and Statista (for 2023, and quarterly) have been considered ahead of the value computed as GDP/Population, because only work force is considered in GDP per capita estimate
	M	Inflation	%	quarterly	World Bank Open Data	Computed from GDP and Real GDP through GDP Deflator

	M	Recession Indicator	binary	ad-hoc	n/a computed binary variable	1 if GDP decreases for three consecutive prior quarters, 0 otherwise
	I	Transport Energy and Environmental taxes	USD million	annual	Statista	Used as a proxy for number of environmental taxes in Korea during the period of the permit regulation (2023 interpolated using linear fit)
	E	Extreme temperature indicator	#	ad-hoc	IEA - Climate Resilience Policy Indicator	Measured as the sum of heating and cooling days for the year, and extrapolated from 2021-2023 using the linear regression fit. Backfilled within the year from full-year estimates
	E	Average Global Temperature Anomaly in Celsius	degC	monthly	NOAA	Backfilled from available values (Local Korea surface temp from IEA not considered)
	T	Cumulative VCM market surplus by sector (Issuance minus retirement) tCO2e	tCO2e	monthly	Quantum Commodity Intelligence	Backfilled from available values

4. EU-ETS allowance price prediction - Feature Set

	Factor type	Variable	Units	Frequency	Source(s)	Data Management
Target (Y)	n/a	Allowance price	USD / ton	daily	ICAP Allowance price explorer	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
Feature (Xi)	S	Compliance Period Identifier	n/a	various	European Commission	'Third' or 'Fourth' based on dates. This is also encoded as its own categorical variable.
	I	Market Stability Action	tCO2e	ad-hoc	EU-ETS Policy Action and MRV reports; Report on the functioning of the european carbon market	Indicator to show if reserve was utilized to regulate price or emissions. Denoted as amount of allowances bought or sold
	I	Number of new free Allowances Introduced	#	annual	EU-ETS Free Allocation reports	Forward fill, followed by backfill to make into continuous series
	I	Number of new auctioned allowances	#	semi-annual	EU-ETS Auctioning Reports	Forward filled between auction dates to make into continuous series
	I	Weighted average sectoral abatement difficulty	Index (0,1)	ad-hoc	Constructed from Reports on functioning, MRV reports and ICAP Factsheet, based on sectoral inclusion	An index between 0 and 1 prepared by weighting sectoral additions using a sector weights dictionary
	I	Daily Traded Volume	#	daily	European Union Transaction Log (EUTL)	Pandas linear interpolation for the unavailable dates
	I	Cap	tCO2e	annual	EU-ETS Cap and Allowances, and ICAP Factsheet	Forward filled; Cap is set in advance, but carried for the entire year
	I	Annual Cap Reduction	tCO2e	annual	EU-ETS Cap and Allowances, and ICAP Factsheet	Cap (previous year) - Cap (current year)
	I	Penalty for	USD /	ad-hoc	ICAP Factsheet	allowance price+108.13

		Non-Compliance	ton			
	I	Regime Name	n/a	ad-hoc	Wikipedia - France, Germany, Poland, EU Parliament, EU Commission, EU Council, Turkey, Netherlands	One field per considered EU country (France, Germany, Poland, EU Parliament, EU Commission, EU Council, Turkey, Netherlands). Names of the Heads of State between the dates they served in the respective nation
	I	Regime Type	Index (0.1)	daily	Wikipedia - France, Germany, Poland, EU Parliament, EU Commission, EU Council, Turkey, Netherlands	Score assigned to each regime based on their political party affiliation. 0 - Right Wing, 0.5 - Centrist and 1 - Left Wing. Overall regime type is calculated as the population weighted regime type index between 0 and 1
	M	Population	#	annual	FRED	One field per considered EU country (France, Germany, Poland, EU Parliament, EU Commission, EU Council, Turkey, Netherlands) and summed. Backfilled from annual reported to daily frequency
	M	GDP	USD	quarterly	FRED	One field per considered EU country (France, Germany, Poland, Turkiye, Netherlands) and averaged for the EU. Backfilled from quarterly reported to daily frequency
	M	GDP/Capita	USD	annual	WorldBank Open Data	GDP/Capita for the European Union. Backfilled from annual values reported at end of year
	M	Inflation	%	quarterly	WorldBank Open Data	Backfilled from reported annual values
	M	Recession Indicator	binary	ad-hoc	n/a copmuted binary variable	1 if GDP decreases for three consecutive prior quarters, 0 otherwise
	I	Revenues from Environmental Taxes	USD	annual	EU Commission, Eurostat	Revenues from all environmental taxes (incl. ETS revenues) across all EU member states is used as a proxy, since there are many environmental taxes. (2023 is projected using linear regression fit)
	E	Economic losses from extreme weather events	USD	annual	EEA (1, 2) and Millman	Economic loss from extreme events is used as a proxy for count of extreme events; Backfilled from reported annual numbers. 2023 is forecasted using 3 year linear-reg fit
	E	Average Global Temperature Anomaly in Celsius	degC	monthly	NOAA	Backfilled from available values
	T	Cumulative VCM market surplus by sector (Issuance minus retirement) tCO2e	tCO2e	monthly	Quantum Commodity Intelligence	Backfilled from available values

5. China ETS (Beijing + Shanghai) allowance price prediction - Feature Set

	Factor type	Variable	Units	Frequency	Source(s)	Data Management
Target (Y)	n/a	Allowance price	USD / ton	daily	ICAP Allowance price explorer	Average of Beijing ETS and Shanghai ETS prices
	n/a	Beijing ETS Allowance price	USD / ton	daily	ICAP Allowance price explorer	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
	n/a	Shanghai ETS Allowance price	USD / ton	daily	ICAP Allowance price explorer	Discontinuity in series handled through cubic spline interpolation, forward filling end dates, backfilling early dates and removing market holidays (non-traded days)
Feature (Xi)	S	Compliance Period Identifier	n/a	various	ICAP Factsheet for Beijing and Shanghai	One per ETS; 2017-2023 (year-wise) for Beijing, 'Phase Two' for Shanghai. Finally concatenated into a string for China as a whole. This is also encoded as its own categorical variable.
	I	Market Stability Action	n/a	various	Beijing Municipal Ecology and Environment Bureau and Shanghai Municipal Bureau of Ecology and Environment	Indicator to show if reserve was utilized to regulate price or emissions. Denoted as amount of allowances bought or sold
	I	Number of new free Allowances Introduced	#	annual	Beijing Municipal Ecology and Environment Bureau and Shanghai Municipal Bureau of Ecology and Environment	In the absence of allotment data from BJ or SH ETS exchanges, or academia has been considered 95% of cap (for Beijing) or 97% of cap (for Shanghai)
	I	Number of new auctioned allowances	#	semi-annual	ICAP Factsheet for Beijing and Shanghai	In the absence of allotment data from BJ or SH ETS exchanges, or academia has been considered 5% of cap (for Beijing) or 3% of cap (for Shanghai)
	I	Weighted average sectoral abatement difficulty	Index (0,1)	ad-hoc	ICAP Factsheet for Beijing and Shanghai	An index between 0 and 1 prepared by weighting sectoral additions using a sector weights dictionary
	I	Traded Volume	tCO2e	annual	Lyu (2021) and estimated from 2020 onwards using linear-fit and actuals for Beijing (BEA) from bjets.com.cn	Pandas linear interpolation for the unavailable dates
	I	Cap Beijing	tCO2e	annual	Beijing Municipal Ecology and Environment Bureau	Forward filled; Cap is set in advance, but carried for the entire year
	I	Cap Shanghai	tCO2e	annual	Shanghai Municipal Bureau of Ecology and Environment	Forward filled; Cap is set in advance, but carried for the entire year
	I	Cap	tCO2e	annual	Beijing Municipal Ecology and Environment Bureau and Shanghai Municipal Bureau of Ecology and Environment	Sum of the Beijing and Shanghai ETS Caps
	I	Annual Cap Reduction	tCO2e	annual	Beijing Municipal Ecology and Environment Bureau	Cap (previous year) - Cap (current year)

	I	Penalty for Non-Compliance	USD	ad-hoc	ICAP Factsheet for Beijing and Shanghai	Max(5x trailing 6mo average price + USD 7,058) for Beijing, (1x allowance price + USD 14,116) for Shanghai)
	I	Regime Name	n/a	ad-hoc	Wikipedia	Names of the CCP Chairman between the dates they served
	I	Regime Type	Index (0.1)	daily	n/a	Assigned as 0.5 as the CCP is considered a centrist party
	M	Population	#	annual	World Bank and focus-economics.com	Backfilled from annual reported to daily frequency
	M	GDP	USD	quarterly	FRED (MKTGDPCNA646NWDB)	Backfilled from quarterly reported to daily frequency
	M	GDP/Capita	USD	quarterly	World Bank Open Data and statista	The values from WBOpen Dat and Statista (for 2023, and quarterly) have been considered ahead of the value computed as GDP/Population, because only work force is considered in GDP per capita estimate
	M	Inflation	%	quarterly	World Bank Open Data	Backfilled from annual reported values
	M	Recession Indicator	binary	ad-hoc	n/a computed binary variable	1 if GDP decreases for three consecutive prior quarters, 0 otherwise
	I	Other Carbon tax	binary	ad-hoc	OECD	If there was any other prevailing carbon taxes in China during the period of the permit regulation
	E	Beijing Air Quality Index	mg/m3	annual	Statista	Beijing's air quality is typically the most reported climate hazard in China (as air quality is linked to extreme temperatures), and is hence used as a proxy for extreme climate events.
	E	Average Global Temperature Anomaly in Celsius	degC	monthly	NOAA	Backfilled from available values
	T	Cumulative VCM market surplus by sector (Issuance minus retirement) tCO2e	tCO2e	monthly	Quantum Commodity Intelligence	Backfilled from available values

APPENDIX D1: CODE FOR CEEMDAN-LSTM

```
def CL_MODEL(data, cutoff):
  data = pd.Series(data['Y'])
  df_ceemdan = cl.decom(data)
  df_ceemdan.plot(title='CEEMDAN Decomposition', subplots=True, figsize=(6,
1*(df_ceemdan.columns.size))) # plot
  plt.show()

  model = Sequential()
  model.add(LSTM(250, input_shape=(2556, 9), activation='tanh'))
  model.add(Dropout(0.5))
  model.add(Dense(1,activation='tanh'))
  model.compile(loss='mse', optimizer='adam')
  kr = cl.keras_predictor(KERAS_MODE=model, FORECAST_HORIZONS=cutoff,
FORECAST_LENGTH=cutoff)

  print('\n1.Sample Entropy Calculate')
  print('-------------------------------')
  df_sampen = cl.inte_sampen(df_ceemdan)
  df_sampen.plot(title='Sample Entropy') # plot
  plt.show() # plot

  print('\n2.K-Means Cluster by Sample Entropy')
  print('-------------------------------')
  df_integrate_form = cl.inte_kmeans(df_sampen)
  print(df_integrate_form) # show

  print('\n3.Integrate IMFs')
  print('-------------------------------')
  df_integrate_result = cl.inte(df_ceemdan, df_integrate_form)
  df_integrate_result = df_integrate_result[0]
  df_integrate_result.plot(title='Integrated IMFs (Co-IMFs) of CEEMDAN', subplots=True,
figsize=(6,3)) # plot
  plt.show() # plot

  print('\n4.Predict Co-IMF0 thru Co-IMF8 by vector-input LSTM (respective method)')
  series_add_predict_result=pd.DataFrame()
  for IMF in df_integrate_result.drop(['target'],axis=1):
    raw, evaluation, train_loss = kr.keras_predict(df_integrate_result[str(IMF)])
    print('======{} Predicting Finished======\n'.format(str(IMF)), evaluation) # show
    raw.plot(title='{} Predicting Result'.format(str(IMF))) # plot
    train_loss.plot(title='{} Training Loss'.format(str(IMF))) # plot
    series_add_predict_result[str(IMF)]=raw['predict']
  series_add_predict_result=series_add_predict_result.sum(axis=1)
  forecast_length = len(raw)
  df_add_predict_raw = pd.DataFrame({'predict': series_add_predict_result.values, 'raw':
data[-forecast_length:].values}, index=range(forecast_length))
  df_add_evaluation = cl.eval_result(data[-forecast_length:],series_add_predict_result)

  print('\n9. Add the result to get the final forecasting results (30 days)')
  print('-------------------------------')
  print('======Hybrid CEEMDAN-LSTM Keras Forecasting Finished======\n',
df_add_evaluation) # show
  df_add_predict_raw.plot(title='Hybrid CEEMDAN-LSTM Keras Forecasting Result') # plot
  plt.show() # plot

  # Evaluation Stats
  print("Mean Absolute Error : " +
str(mean_absolute_error(data[-forecast_length:],series_add_predict_result)))
  print("Root Mean Squared Error : " +
str(sqrt(mean_squared_error(data[-forecast_length:],series_add_predict_result))))
  print("Mean Absolute Percentage Error : " +
str(mean_absolute_percentage_error(data[-forecast_length:],series_add_predict_result)))
```

APPENDIX D2: CODE FOR XGBOOST

```
def XGraph(modelXG, X_unseen, Y_unseen):
  y_pred = modelXG.predict(X_unseen)
  y_pred = pd.DataFrame(y_pred,index=Y_unseen.index)
  # Plot the results
  plt.plot(y_pred)
  plt.plot(Y_unseen)
  plt.xlabel("Days")
  plt.ylabel("Output")
  plt.legend(["Model","Actual"])
  plt.gcf().autofmt_xdate()
  plt.show()
def XGBOOST_MODEL(X_train, Y_train, X_test, Y_test, X_unseen, Y_unseen):
  parameters = {
    'n_estimators': [100, 150, 200, 250, 300],
    'learning_rate': [0.005, 0.01, 0.05, 0.1],
    'max_depth': [6, 8, 10],
    'gamma': [0.001, 0.005, 0.01, 0.02],
    'random_state': [42]
  }
  eval_set = [(X_train, Y_train), (X_test, Y_test)]
  model = xgb.XGBRegressor()
  clf = GridSearchCV(model, parameters)
  clf.fit(X_train, Y_train)

  print(f'Best params: {clf.best_params_}')
  print(f'Best validation score = {clf.best_score_}')

  # Use the best parameters and fit the model
  model = xgb.XGBRegressor(**clf.best_params_, early_stopping = 5,
objective='reg:squarederror')
  model.fit(X_train, Y_train, eval_set= [(X_test, Y_test)], verbose=False)

  # Evaluation Stats
  predictions = model.predict(X_test)
  print("Mean Absolute Error : " + str(mean_absolute_error(Y_test, predictions)))
  print("Root Mean Squared Error : " + str(sqrt(mean_squared_error(Y_test,
predictions))))
  print("Mean Absolute Percentage Error : " + str(mean_absolute_percentage_error(Y_test,
predictions)))

  # Graph the model with unseen data
  XGraph(model, X_unseen, Y_unseen)
```

APPENDIX D3: CODE FOR TRANSFORMER

```
def TransformerTensors(data):
  return torch.tensor(data.values, dtype=torch.float32)

def transformer_model(input_shape=(14,), d_model=14):
    inputs = layers.Input(shape=input_shape)

    # Positional encoding
    position = tf.range(start=0, limit=input_shape[0], delta=1)
    position = layers.Embedding(input_shape[0], d_model)(position)

    # Transformer encoder layers
    x = layers.Dense(200, activation='relu')(inputs)
    x = layers.LayerNormalization()(x)
    x = layers.Dense(200, activation='relu')(x)
    x = layers.LayerNormalization()(x)
    x = layers.Dense(200, activation='relu')(x)
    x = layers.LayerNormalization()(x)

    # Output layer
    outputs = layers.Dense(1, activation='linear')(x)

    model = Model(inputs=inputs, outputs=outputs)
    return model

def Transformer_Eval(X_Train, Y_Train):
  model = transformer_model()
  model.compile(optimizer='adam', loss='mean_squared_error')

  # Train the model
  model.fit(X_Train, Y_Train, epochs=20, batch_size=100, validation_split=0.2)

  # Evaluation Stats
  predictions = model.predict(X_test)
  print("Mean Absolute Error : " + str(mean_absolute_error(Y_test, predictions)))
  print("Root Mean Squared Error : " + str(sqrt(mean_squared_error(Y_test,
predictions))))
  print("Mean Absolute Percentage Error : " + str(mean_absolute_percentage_error(Y_test,
predictions)))

  # Predict the unseen data
  with torch.no_grad():
    y_pred = model(X_unseen.values)
  y_pred = pd.DataFrame(y_pred,index=Y_unseen.index)
  # print(y_pred)
  # Plot the results
  plt.plot(y_pred)
  plt.plot(Y_unseen)
  plt.gcf().autofmt_xdate()
  plt.xlabel("Input")
  plt.ylabel("Output")
  plt.legend(["Model","Actual"])
  plt.show()
```

APPENDIX E: RESULTS DETAILED

		RGGI	California	Korea	EU-ETS	China
CEEMDAN-LSTM	**MAE**	0.5036	0.2579	0.3527	0.5948	0.4148
	RMSE	0.5295	0.3090	0.3732	0.7632	0.4457
	MAPE	0.0310	0.0080	0.0512	0.0075	0.0330
XGBoost	**MAE**	1.6112	5.8878	5.6588	20.3251	3.3277
	RMSE	2.2885	6.3637	6.3357	23.7090	3.8384
	MAPE	0.1179	0.1935	0.3191	0.2269	0.3464
Transformer	**MAE**	8.8278	13.3092	6.8605	86.6462	4.9140
	RMSE	8.8869	13.4667	7.5312	87.2383	5.2179
	MAPE	0.6537	0.4433	0.3985	1.0132	0.4969

APPENDIX F: LITERATURE REVIEW - ENVIRONMENTAL ETF DESIGN

ETF	[KraneShares Global Carbon Strategy ETF]	[HANetf SparkChange Physical Carbon EUA ETC]	[Global X Carbon Credits Strategy ETF]
Issuer or Proponent	Krane Funds Advisors, LLC and Climate Finance Partners LLC	HANetf ETC Securities plc	Global X Management Company LLC
Portfolio Manager	James Maund	Nik Bienkowski and Elliot Waxman	2008
Market Cap / Net Assets	USD 303.75 million	USD 169.76 million	USD 1.35 million (liquidated)
Theme	Investing in 4 major ETS between US and Europe	Focuses on decarbonizing companies only	Variable weight ETF across RGGI, CA, EU, UK
Launched	Jul 2020	Oct 2021	May 2023
Ticker	KRBN	CO2.L, CO2U.L ($), CO2P.L (£) FCO2 GY, CO2 IM	NTRL CUSIP: 37960A586
Exchange	NYSE Arca	LSE, XETRA, Borsa Italiana	NYSE Arca
Goal	To provide a total return (before fees/expenses) tracks IHS Markit Global Carbon Index (the “Index”) (volume weighted index)	Prevent carbon emissions while providing access to one of the best performing commodities, creating measurable environmental impact	To provide investment results that correspond generally to the price and yield performance, before fees/expenses of the ICE Global Carbon Futures Index
Markets	EU-ETS, California, RGGI, UK ETS	EU-ETS (EUAs)	EU-ETS, California, RGGI, UK ETS
Underlyings	Carbon Futures (physical delivery), Forwards Debt Instruments (short-term bond funds, bond etfs, treasury, corporate <=12m) Currency futures (hedge)	Physical EUAs	Carbon Futures (physical delivery) (varying weights)
Specifications	Max 2y vintage Trading volume >= $10m Min weight 5% Max weight 65%	Entitlement per security: 0.978	EUA Futures
Currency	USD	EUR	USD
Legal Structure	Fund + Subsidiary (25% invested by fund)	Issuer (Ireland) + Backing Issuer (Jersey)	Global X (advisor) + Global X Funds (subsidiary) + Mirae Asset Financial Group (Fund Administration)
Domicile	Cayman Islands	Ireland	United States
Applicable Law	Cayman Islands, U.S. Securities	UK, EU MiFID2	United States
Regulators	SEC, CFTC	Financial Conduct Authority	SEC
Investor Qualification	Market Makers / Broker-Dealers	Experienced Financial Institutions	Qualified institutional investors
Minimum Investment	50,000 shares	1 Carbon security	I share
Return	108.83% (2021) -10.51% (2022) 6.65% (2023)	CAGR 55% since 2018 6m: -26.13% 12m: -33.38%	-15.57% (May 2023-Apr 2024)
Risks	Carbon Market Failure Regulatory Subsidiary Investment & Tax Derivatives Risk Commodity Derivatives Risk Strategic & Management Futures ‘Roll’ Risk Commodity Pool Registration Geographic Focus EU & California Fixed Income & Currency U.S. G.O. & Cash-equivalent Concentration Risk Secondary Market Liquidity Risk Market Risk Tracking Error Turnover and Brokerage Large Shareholder	Underlying Allowance Regimes Value of Carbon Securities Geopolitical & Macroeconomic Market Risk Operational Risks Security of Secured Property and Backing Issuer Assets Issuer Legal Structure Backing Issuer Legal Structure Legal Risks Technological Slowdown Policy & Government Regime Inflation, Rates and Currency Secondary Market Liquidity Cybersecurity Risks Ireland & Jersey Risk	Carbon Market Failure Regulatory Subsidiary Investment & Tax Derivatives Risk Commodity Derivatives Risk Strategic & Management Leverage Futures ‘Roll’ Risk EU & California Fixed Income & Currency Concentration Risk Secondary Market Liquidity Risk Market Risk Turnover and Brokerage Large Shareholder
Fees and expenses	Management 0.78% Other expenses 0.01% Transaction costs Taxes	Management Fee Redemption Fee Total expense ratio (0.89 p.a.) Transaction costs Taxes	Expense Ratio (0.39%) Management Fee Transaction costs Taxes

ETF	Barclays iPath $80 million Series B Carbon ETN	China International Capital Corporation, Hong Kong Asset Management Limited, Carbon Futures ETF	WisdomTree Classic Commodity Securities and Longer Dated Commodity Securities ETF
Issuer or Proponent	Barclays Bank PLC	CICC HKAM Limited	WisdomTree Commodity Securities Limited
Portfolio Manager	n/a	Brown Brothers Harriman Trustee Services (Hong Kong)	Bryan Governey, Christopher Foulds, Steven Ross, Peter M Ziemba
Market Cap / Net Assets	USD 25.39 million	USD 62.11 million	USD 145.227 million
Theme	Focuses on futures contracts in EUA and CER	Futures contracts in the Chinese/Hong Kong market	Focused on raw material and commodities
Launched	Sep 9, 2019	Mar 2022	Jun 2017
Ticker	GRN (ISIN: US06747C3227)	03060 (HKD), 83060 (RMB), 09060 (USD)	CARB (LSE)
Exchange	NYSE Arca	HKSE (3060.HK)	Euronext Dublin // LSE
Goal	To provide exposure to the price of carbon as measured by the return of ICE carbon futures contracts and to track the Barclays Global Carbon II TR USD Index	Tracker Fund: To provide investment results that, before fees and expenses, closely correspond to the performance of the ICE EUA Carbon Futures Index	To provide a carbon security within the WisdomTree series of commodity funds which replicates SOLCARBT (Solactive Carbon Emission Rolling Futures Excess Return Index)
Markets	Heavily EU-ETS weighted	EU-ETS	EU-ETS
Underlyings	ICE Carbon Futures + Some Credits from Kyoto Protocol CDM program (<0.1%)	ICE EUA Carbon Futures (excess return index)	ICE EUA Carbon Futures (physical delivery)
Specifications	UDS 8m ETN outstanding at $10 per ETN; Only pays return at maturity. Money market instrument.	Full index replication strategy with same weights (with some discretion)	n/a
Currency	USD	HKD / RMB / USD	EUR (translated to USD)
Legal Structure	Principal (Barclays Bank PLC) + Affiliate (Barclays Capital Inc.)	Fund manager (CICC Funds Series) + Sub-fund (CICC Carbon Futures ETF Subsidiary)	Fund + Subsidiary + Trustee
Domicile	United States	Hong Kong	Ireland / UK
Applicable Law	U.S. Securities	Hong Kong Securities Law	EU MiFID II, Jersey
Regulators	SEC, FCA	HKMA, SFC (Securities and Futures Commission, HK)	EU Parliament, Central Bank of Ireland, FCA, Jersey FSC
Investor Qualification	Market Makers / Brokerages	Market Makers / Broker-Dealers	Market Makers / Broker-Dealers
Minimum Investment	1 ETN ($10)	I share = 1 EUA Future = 1,000 carbon emission allowances	1 commodity security
Return	-2.99% (2023), -0.07% (2022) 147.21% (2021), -32.93% (1Y) 12.59% (3Y), 10.73% (1m)	-8.92% (2023) Tracking error (-1.05%)	-25.01% (1Y) -15.65% (2024 YTD)
Risks	Carbon Market Disruption or Suspension, Regulatory Market Risk Uncertain Principal Repayment Credit or Issuer Risk Issuer Redemption (call) Risk Lack of Diversification (EU risk) No Interest Payments Illiquid Secondary Market Tracking Error, Futures 'Roll' Risk Subsidiary Investment & Tax Strategic & Management Turnover and Brokerage Currency Risk Settlement Risk Index Force Majeure Event	Carbon Market Failure Carbon Futures Market Risk FX Derivatives Risk (hedge) Regulatory Risk Subsidiary Investment & Tax Energy Sector Risk Liquidity Risk Contango and Backwardation Margin Risk Spot-Futures correlation drift Roll Risk No Principal Guarantee Tracking error	Commodity prices & vol Carbon allowance price & vol Carbon Market Failure Regulatory Subsidiary Investment & Tax Derivatives Risk Strategic & Management Futures 'Roll' Risk No Principal Guarantee Tracking error Geopolitical Risk Early Redemption (Call) Risk Liquidity & Market Risk Counterparty and Issuer Risk Calculation Agent Conflict
Fees and expenses	Investor Fee 0.75% Transaction costs Taxes	Brokerage Transaction Levy - 0.0027% Trading Fee - 0.005% Inter-counter transfer - HKD5 / instruction Net Management Fee 0.99% Taxes	Management & License Fee (0.35 - 0.49% pa); Application Fee ((0.08%), Commodity Contract Counterparty Fee Calculation Agent Fee Transaction costs, Taxes

ETF	Harbor Energy Transition Strategy ETF	Korea Retail Carbon ETF	World Carbon Fund
Issuer or Proponent	Harbor Capital	Local Korean Securities Firm	Carbon Cap Management LLP
Portfolio Manager	Matthew Schwab (Quantix Commodities LP)	n/a	Alex Child and Tor Sinclair
Market Cap / Net Assets	USD 20.24 million	n/a	USD 190.58 million
Theme	Focus on commodities to assist in the transition	First Asian ETF carbon product	Invest in liquid and regulated carbon markets using Core Long + Hedge, and Alpha Strategies
Launched	Jul 2022	Mar 2024	Feb 2020
Ticker	RENW	n/a	n/a
Exchange	NYSE Arca	Korea Exchange	Private
Goal	to provide the opportunity to invest in this transition with the commodities needed to facilitate change as the world marches towards a net zero carbon emissions goal	First Asian Retail ETF in note form (ETN)	absolute return strategy, seeking to deliver positive returns with a low correlation to traditional and alternative asset classes as well as a direct impact on climate change
Markets	EU-ETS and California	K-ETS	EU-ETS, WCI, RGGI, New Zealand ETS
Underlyings	Tracks the Quantix Energy Transition Index (QET)	Carbon Futures or Spot Allowances	Carbon Futures
Specifications	QET includes the commodities copper, aluminum, nickel, zinc, lead, natural gas, silver, palladium, platinum, soybean oil, ethanol, and EUA / California CCA (2% min weight, 15% max weight)	n/a	20% of the performance fees are committed to buying and permanently retiring carbon allowances
Currency	USD	USD / KRW	EUR, USD and GBP
Legal Structure	Offered as a 1940-Act ETF structure, avoiding the need for Schedule K-1 tax filing	n/a	"Article 9" fund under the EU's Sustainable Finance Disclosure Regulation (SFDR), Irish QIAIF Structure
Domicile	United States	Korea	Ireland
Applicable Law	U.S. Securities	Korea Securities	EU SFDR, UK
Regulators	SEC, CFTC	Ministry of Environment Financial Services Commission (FSC)	EU MiFID II, FCA
Investor Qualification	Market Makers / Broker-Dealers	Retail investors via Brokerage Firms	Institutional or Private Investors
Minimum Investment	1 share	n/a	Class I: $5 million Class P: $250,000
Return	-26.91% (1Y) -30.71% (2023 total return) -7.09% (2024)	n/a	52.60% (2021), 8.96% (2022), 14.60% (2023),
Risks	Commodity and Commodity Linked Derivatives Risk Concentration Risk Authorized Participant Concentration / Trading Risk Energy Transition Risk Foreign Currency Risk Foreside Fund Services, LLC	Typical risk factors affecting ETFs and ETNs in the carbon markets	Sharpe Ratio (12m) = 1.05 Max. Drawdown (12m) = -3.7% Typical qualitative risk factors affecting ETFs in the carbon markets
Fees and expenses	Total Expense Ratio 0.80% Transaction costs, Taxes	n/a	Class I: 1.5% Mgmt, 15% Performance Class P: 2% Mgmt, 20% Performance

APPENDIX G: CLIMATE SCENARIOS

IPCC (Intergovernmental Panel on Climate Change), AR6 (Sixth Assessment Report)) Scenarios **Source: [IPCC (Page 1810)](), [USDA](), [Carbon Brief]()**		
Scenario ID	*Scenario Name*	*Scenario description*
(Note: Radiative forcing (downward - upward radiative flux) is a concept used to quantify the change in energy balance in the Earth's atmosphere and is impacted by factor such as GHG concentrations, aerosols, changes in surface albedo and solar irradiance)		
Representative Concentration Pathways (RCP) from CMIP5 model *Future greenhouse gas concentrations (2000-2100) in ppm under different radiative forcing*		
RCP 1.9 **Paris**	RCP 1.9 W/m^2 radiative forcing value in 2100	● Radiative forcing value of 1.9 W/m^2 in 2100 ● Attainable only with stringent and coordinated mitigation efforts of all Paris Agreement Member States ● Temperature rise (from global warming) by 2100: <1.5 ℃
RCP 2.6 **Very stringent**	RCP 2.6 W/m^2 radiative forcing value in 2100	● Radiative forcing peaks at 3 W/m^2 and declines to 2.6 W/m^2 in 2100. ● Attainable only with stringent mitigation efforts such as CO_2 emission reductions post 2020 ● Temperature rise (from global warming) by 2100: <2 ℃
RCP 3.4 **Intermediate pathway**	RCP 3.4 W/m^2 radiative forcing value in 2100	● Radiative forcing value of 3.4 W/m^2 in 2100 ● Attainable with stringent mitigation efforts such considerable atmospheric GHG reduction ● Temperature rise (from global warming) by 2100: <2-2.4 ℃
RCP 4.5 **Intermediate scenario**	RCP 4.5 W/m^2 radiative forcing value in 2100	● Future greenhouse gas concentrations (2000-2100) in ppm based on the attainment of a radiative forcing value of 4.5 W/m^2 in 2100 ● Realistic baseline factoring exhaustibility of fossil-fuels ● Attainable only with reasonable mitigation efforts such as halving of 2050 CO_2 levels by 2100 ● Temperature rise (from global warming) by 2100: <2-3 ℃
RCP 6 **Stabilization**	RCP 6 W/m^2 radiative forcing value in 2100	● Radiative forcing value of 6 W/m^2 in 2100 ● Emissions peak around 2080 and then decline ● Attainable only with technology based mitigation efforts amidst high GHG emissions rate ● Temperature rise (from global warming) by 2100: <3-4 ℃
RCP 7 **Baseline outcome**	RCP 7 W/m^2 radiative forcing value in 2100	● Radiative forcing value of 7 W/m^2 in 2100 ● Considered the plausible baseline scenario outcome and corresponds to SSP 3 (Regional Rivalry)
RCP 8.5 **High**	RCP 8.5 W/m^2 radiative forcing value in 2100	● Future greenhouse gas concentrations (2000-2100) in ppm based on the attainment of a radiative forcing value of 4.5 W/m^2 in 2100. ● Higher GHG emissions, hence higher global surface temperatures and greater effects of climate change. ● Temperature rise (from global warming) by 2100: <4 ℃
Shared Socioeconomic Pathways (SSP) from CMIP6 model *Future greenhouse gas concentrations (2000-2100) in ppm under different projections of socioeconomic global change and climate policies based on underlying factors such as population, technology, economic growth and education levels*		
SSP 1 **390 ppm (2100)**	Sustainability (≡ RCP 1.9 and RCP2.6)	"The world shifts gradually, but pervasively, toward a more sustainable path, emphasizing more inclusive development that respects perceived environmental boundaries."
SSP 2	Middle of the Road (≡ RCP 4.5)	"The world follows a path in which social, economic, and technological trends do not shift markedly from historical patterns."
SSP 3	Regional Rivalry (≡ RCP 7)	"A resurgent nationalism, concerns about competitiveness and security, and regional conflicts push countries to increasingly focus on domestic or, at most, regional issues."
SSP 4	Inequality (≡ RCP 3.4)	"Highly unequal investments in human capital, combined with increasing disparities in economic opportunity and political power, lead to increasing inequalities and stratification both across and within countries."
SSP 5 **1130 ppm (2100)**	Fossil-fueled Development (≡ RCP 8.5)	"This world places increasing faith in competitive markets, innovation and participatory societies to produce rapid technological progress and development of human capital as the path to sustainable development. Global markets are increasingly integrated."
IEA (International Energy Agency) Scenarios **Source: [IEA]()**		
Scenario ID	*Scenario Name*	*Scenario description*
NZE *(Introduced: 2023)*	Net Zero Emissions by 2050 Scenario	● Normative scenario for Net Zero emissions by 2050 also attaining other targets such as universal energy and clean cooking access by 2030 (SDGs), full CO_2 offsets and global temperature increase (>1.5 ℃) ● Attainment is based on a portfolio of clean energy technologies, orderly energy transition minimizing energy volatility and stranded assets with international cooperation ● Requires enhanced financial support to emerging markets and developing economies
APS	Announced Pledges	● Scenario where countries implement their national targets (2030 NDC, announced or legislative

(Introduced: 2021)	Scenario	targets) and other 2050 (longer-term) targets which may not attain universal energy and clean cooking access (2030), full CO2 offsets or limit global temperature increases to <1.5 ℃ ● Attainment is based on a new technologies with limits to technological transfers, reasonably orderly energy transition and limited international cooperation ● Assumes full attainment of self-determined goals by countries, and is designed to demonstrate how current pledges help the world get close to the 1.5 ℃ temperature increases ● The gap between NZE and APS scenarios is called the Paris Agreement 'ambition gap'
STEPS *(Includes policies upto: Aug 2023)*	Stated Policies Scenario	● Scenario built based on a detailed (bottom-up), sector-by-sector review of policies (active and under development) which have been implemented to attain stated goals (2030 NDC, announced or legislative targets and other 2050 longer-term) ● Considers pricing policies, efficiency standards and schemes, electrification programmes, and specific infrastructure projects ● Attainment is based on realization of policies, a new technologies with limits to technological transfers, managable energy transition, limited international cooperation and industry action to meet capacity for clean energy technologies. ● Conservative and assumes less than full attainment of self-determined goals by countries, and is designed to demonstrate how current pledges help the world get close to the 1.5 ℃ temperature increases ● Still a large gap between STEPS and APS/NZE outcomes for 2050
SDS *(Introduced: Dec 2019)*	Sustainable Development Scenario	● A scenario focussed on full-attainment of the UN SDGs (indicators and targets) by 2030 and its implications on energy sector and global CO2 emissions and global temperature increases from global warming.
NGFS (Network for Greening the Financial System) Scenarios **Source: NGFS**		
Scenario ID	*Scenario Name*	*Scenario description*
ORD	Orderly	● "assume climate policies are introduced early and become gradually more stringent. Both physical and transition risks are relatively subdued"
DIS	Disorderly	● "explore higher transition risk due to policies being delayed or divergent across countries and sectors. Carbon prices are typically higher for a given temperature outcome"
HHW	Hot house world	● "assume that some climate policies are implemented in some jurisdictions, but global efforts are insufficient to halt significant global warming. Critical temperature thresholds are exceeded, leading to severe physical risks and irreversible impacts like sea-level rise"
TLTL	Too little, too late	● "reflect delays and international divergences in climate policy ambition that imply elevated transition risks in some countries and high physical risks in all countries due to the overall ineffectiveness of the transition"

Acknowledgments

The team would like to thank their families, WorldQuant University and our peers from the period Jun 2022 - May 2024, for giving us the opportunity and support to complete this rigorous graduate-level course and to carry out research on this topic of deep interest to the team.

7. REFERENCES

Amanatidis, Georgios. "European policies on climate and energy towards 2020, 2030 and 2050." (2019) https://www.researchgate.net/publication/331207796_European_policies_on_climate_and_energy_towards_2020_2030_and_2050

Anishnama. "Understanding Bidirectional LSTM for Sequential Data Processing." Medium, 18 May 2023, medium.com/@anishnama20/understanding-bidirectional-lstm-for-sequential-data-processing-b83d6283befc

Arango,Isabel, Marshall,George,Conway,Darragh,Webster,Robin,Comeau,Louise,Besley,Daniel James,Saldarriaga. "Guide to Communicating Carbon Pricing." World Bank, documents.worldbank.org/en/publication/documents-reports/documentdetail/668481543351717355/guide-to-communicating-carbon-pricing.

Atsalakis, George S. "Using Computational Intelligence to Forecast Carbon Prices." Applied Soft Computing, vol. 43, June 2016, pp. 107–116, www.sciencedirect.com/science/article/pii/S1568494616300801, https://doi.org/10.1016/j.asoc.2016.02.029. Accessed 4 Jan. 2020.

Barrett, Scott. "Montreal versus Kyoto." Global Public Goods, 29 July 1999, pp. 192–219, https://doi.org/10.1093/0195130529.003.0010.

Benz, Eva, and Stefan Trück. "Modeling the Price Dynamics of CO2 Emission Allowances." Energy Economics, vol. 31, no. 1, Jan. 2009, pp. 4–15, www.sciencedirect.com/science/article/pii/S0140988308000972, https://doi.org/10.1016/j.eneco.2008.07.003.

Berta, Nathalie. "3. A Note on John Dales and the Early History of Emissions Trading: Mixing Standards and Markets for Rights." Cahiers d'Économie Politique, vol. n° 79, no. 1, 26 May 2021, pp. 61–84, https://doi.org/10.3917/cep1.079.0061. Accessed 30 Aug. 2021.

Bland, Greg. "Yfinance Library - a Complete Guide - AlgoTrading101 Blog." Quantitative Trading Ideas and Guides - AlgoTrading101 Blog, 16 June 2020, algotrading101.com/learn/yfinance-guide/.

Bloomberg. "Bloomberg." Bloomberg.com, 2019, www.bloomberg.com/.

Byun, Suk Joon, and Hangjun Cho. "Forecasting Carbon Futures Volatility Using GARCH Models with Energy Volatilities." Energy Economics, vol. 40, Nov. 2013, pp. 207–221, www.sciencedirect.com/science/article/pii/S0140988313001448, https://doi.org/10.1016/j.eneco.2013.06.017.

Carbon Pulse. carbon-pulse.com/.

Chen, T. & Guestrin, C. Xgboost: A scalable tree boosting system. KDD. 785–794 (2016).

Chevallier, Julien. "Carbon Futures and Macroeconomic Risk Factors: A View from the EU ETS." Energy Economics, vol. 31, no. 4, July 2009, pp. 614–625, https://doi.org/10.1016/j.eneco.2009.02.008.

Coase, Ronald H. "The Problem of Social Cost" The Journal of Law & Economics, vol. III, Oct. 1960, pp. 1–44, www.jstor.org/stable/724810.

Company, SEI Investments Distribution, et al. "KRBN." KraneShares, kraneshares.com/krbn/.

Crocker, Thomas. "Some Economics of Air Pollution Control." Natural Resources Journal, vol. 8, no. 2, 1 Apr. 1968, p. 236, digitalrepository.unm.edu/nrj/vol8/iss2/3/. Accessed 7 Apr. 2024.

Crocker, Thomas D., et al. "The Theory of Environmental Policy." Land Economics, vol. 52, no. 2, May 1976, p. 255, https://doi.org/10.2307/3145303.

Dong, Hongwei, et al. "A Multi-Strategy Integration Prediction Model for Carbon Price." Energies, vol. 16, no. 12, 9 June 2023, pp. 4613–4613, https://doi.org/10.3390/en16124613. Accessed 9 Mar. 2024.

European Commission. "EU Emissions Trading System (EU ETS)." European Commission, 2023, climate.ec.europa.eu/eu-action/eu-emissions-trading-system-eu-ets_en.

European Environment Agency. "European Environment Agency's Home Page." European Environment Agency, 2017, www.eea.europa.eu/.

Fan, Xinghua, et al. "Chaotic Characteristic Identification for Carbon Price and an Multi-Layer Perceptron Network Prediction Model." Expert Systems with Applications, vol. 42, no. 8, May 2015, pp. 3945–3952, https://doi.org/10.1016/j.eswa.2014.12.047. Accessed 4 May 2019.

Feng, Mengdan, et al. "Carbon Price Prediction Based on Decomposition Technique and Extreme Gradient Boosting Optimized by the Grey Wolf Optimizer Algorithm." Scientific Reports, vol. 13, no. 1, 27 Oct. 2023, p. 18447, www.nature.com/articles/s41598-023-45524-2, https://doi.org/10.1038/s41598-023-45524-2. Accessed 27 Mar. 2024.

Google Earth Engine. "Google Earth Engine." Google.com, 2019, earthengine.google.com/.

Gray, Alex. "5 Charts That Explain the Paris Climate Agreement." World Economic Forum, 4 Nov. 2016, www.weforum.org/agenda/2016/11/5-charts-that-explain-the-paris-climate-agreement/.

Greenstone, Michael. "The Impacts of Environmental Regulations on Industrial Activity: Evidence from the 1970 and 1977 Clean Air Act Amendments and the Census of Manufactures." Journal of Political Economy, vol. 110, no. 6, Dec. 2002, pp. 1175–1219, https://doi.org/10.1086/342808. Accessed 15 Jan. 2020.

Hachcham, Aymane. "XGBoost: Everything You Need to Know." Neptune.ai, 2023, neptune.ai/blog/xgboost-everything-you-need-to-know.

Harvey, Fiona, and Fiona Harvey Environment editor. "Ban Private Jets to Address Climate Crisis, Says Thomas Piketty." The Guardian, 22 Nov. 2023, www.theguardian.com/environment/2023/nov/22/ban-private-jets-to-address-climate-crisis-says-thomas-piketty.

Hausfather, Zeke. "Explainer: How "Shared Socioeconomic Pathways" Explore Future Climate Change." Carbon Brief, 2018, www.carbonbrief.org/explainer-how-shared-socioeconomic-pathways-explore-future-climate-change/.

https://climate.ec.europa.eu/document/download/a514d3cb-a089-42c2-ab95-785f0191f8df_en?filename=thomson_reuters_point_carbon_en.pdf

Hochreiter, Sepp, and Jürgen Schmidhuber. "Long Short-Term Memory." Neural Computation, vol. 9, no. 8, Nov. 1997, pp. 1735–1780, ieeexplore.ieee.org/abstract/document/6795963, https://doi.org/10.1162/neco.1997.9.8.1735.
International Monetary Fund. "Climate Change Data." Climatedata.imf.org, 2023, climatedata.imf.org/pages/climatechange-data.
Kelly, Mick. "Hubert Horace Lamb (1913-97)." Nature, vol. 388, no. 6645, Aug. 1997, pp. 836–836, https://doi.org/10.1038/42156.
Lu, Hongfang & Ma, Xin & Huang, Kun & Azimi, Mohammadamin. (2020). Carbon trading volume and price forecasting in China using multiple machine learning models. Journal of Cleaner Production. 10.1016/j.jclepro.2019.119386.
Montgomery, W. David. "Markets in Licenses and Efficient Pollution Control Programs." Journal of Economic Theory, vol. 5, no. 3, Dec. 1972, pp. 395–418, https://doi.org/10.1016/0022-0531(72)90049-x.
Nordhaus, William D. "Can We Control Carbon Dioxide? (from 1975)." American Economic Review, vol. 109, no. 6, June 2019, pp. 2015–2035, https://doi.org/10.1257/aer.109.6.2015.
Riahi, Keywan, et al. The Shared Socio-Economic Pathways (SSPs): An Overview SSP Population by IIASA 1 , Urbanization by NCAR 4 , GDP by Dellink et Al., IIASA 1 , PIK 3 ; SSP IAM Scenarios by IIASA. https://unfccc.int/sites/default/files/part1_iiasa_rogelj_ssp_poster.pdf
Roy, Samapriya. "Awesome-Gee-Community-Catalog." Gee-Community-Catalog.org, gee-community-catalog.org/.
Rudnik, Katarzyna, et al. "Forecasting day-ahead carbon price by modelling its determinants using the PCA-based approach." Energies, vol. 15, no. 21, 29 Oct. 2022, p. 8057, https://doi.org/10.3390/en15218057.
Scripps Institution Of Oceanography. "The Keeling Curve." Keelingcurve.ucsd.edu, 2023, keelingcurve.ucsd.edu/.
Shahzad, Umer, et al. "Forecasting Carbon Emissions Future Prices Using the Machine Learning Methods." Annals of Operations Research, 6 Feb. 2023, https://doi.org/10.1007/s10479-023-05188-7.
Shi, Hanxiao, et al. "A CNN-LSTM Based Deep Learning Model with High Accuracy and Robustness for Carbon Price Forecasting: A Case of Shenzhen's Carbon Market in China." Journal of Environmental Management, vol. 352, 14 Feb. 2024, p. 120131, pubmed.ncbi.nlm.nih.gov/38266520/, https://doi.org/10.1016/j.jenvman.2024.120131. Accessed 27 Mar. 2024.
Stern, Nicholas. "Getting the Social Cost of Carbon Right | by Nicholas Stern & Joseph E. Stiglitz." Project Syndicate, 15 Feb. 2021, www.project-syndicate.org/commentary/biden-administration-climate-change-higher-carbon-price-by-nicholas-stern-and-joseph-e-stiglitz-2021-02.
Tietenberg, Tom. "Cap-And-Trade: The Evolution of an Economic Idea." Agricultural and Resource Economics Review, vol. 39, no. 3, Oct. 2010, pp. 359–367, https://doi.org/10.1017/s106828050000736x. Accessed 7 Apr. 2022.
Torres, M. E., Colominas, M. A., Schlotthauer, G. & Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. ICASSP. 4144–4147 (2011).
UNFCCC. "Nationally Determined Contributions (NDCs) | UNFCCC." Unfccc.int, 2019, unfccc.int/process-and-meetings/the-paris-agreement/nationally-determined-contributions-ndcs.
UNFCCC. "What Is the Kyoto Protocol? ." UNFCCC, UNFCCC, 2019, unfccc.int/kyoto_protocol.
UNFCCC. "What Is the Kyoto Protocol? ." UNFCCC, UNFCCC, 2019, unfccc.int/kyoto_protocol.
UNFCCC. "The Kyoto Protocol Mechanisms". UNFCCC, UNFCCC, 2007, https://unfccc.int/resource/docs/publications/mechanisms.pdf
United Nations. "The Paris Agreement." United Nations Climate Change, United Nations, 2015, unfccc.int/process-and-meetings/the-paris-agreement.
Wang, Feng, et al. "Carbon Trading Price Forecasting: Based on Improved Deep Learning Method." Procedia Computer Science, vol. 214, 2022, pp. 845–850, https://doi.org/10.1016/j.procs.2022.11.250. Accessed 19 Mar. 2023.
Wang, Yanfeng, et al. "A Novel Deep Learning Carbon Price Short-Term Prediction Model with Dual-Stage Attention Mechanism." Applied Energy, vol. 347, 1 Oct. 2023, p. 121380, ui.adsabs.harvard.edu/abs/2023ApEn..34721380W/abstract, https://doi.org/10.1016/j.apenergy.2023.121380. Accessed 27 Mar. 2024.
Waterville, Colby College. "The Evolution of Emissions Trading by Tom Tietenberg Mitchell Family Professor of Economics." (2010).
Wikipedia Contributors. "An Essay on the Principle of Population." Wikipedia, Wikimedia Foundation, 2 Dec. 2018, en.wikipedia.org/wiki/An_Essay_on_the_Principle_of_Population
Wikipedia Contributors. "An Essay on the Principle of Population." Wikipedia, Wikimedia Foundation, 2 Dec. 2018, en.wikipedia.org/wiki/An_Essay_on_the_Principle_of_Population.
Wisdomtree.eu, 2024, www.wisdomtree.eu/en-gb/pricing/wisdomtree-commodity-securities-limited-classic. Accessed 19 Mar. 2024.
World Bank. "Carbon Pricing Dashboard | Up-To-Date Overview of Carbon Pricing Initiatives." Carbonpricingdashboard.worldbank.org, 2023, carbonpricingdashboard.worldbank.org/.
World Bank. "Carbon Pricing Dashboard | Up-To-Date Overview of Carbon Pricing Initiatives." Worldbank.org, 31 Mar. 2023, carbonpricingdashboard.worldbank.org/map_data.
World Bank. "CO2 Emissions (Metric Tons per Capita) | Data." Worldbank.org, 2020, data.worldbank.org/indicator/EN.ATM.CO2E.PC.
Ye, Peng, et al. "Forecasting the Return of Carbon Price in the Chinese Market Based on an Improved Stacking Ensemble Algorithm." Energies, vol. 16, no. 11, 4 June 2023, pp. 4520–4520, https://doi.org/10.3390/en16114520. Accessed 7 Nov. 2023.

Yun, Po, et al. "Forecasting Carbon Dioxide Emission Price Using a Novel Mode Decomposition Machine Learning Hybrid Model of CEEMDAN‐LSTM." Energy Science & Engineering, 18 Sept. 2022, https://doi.org/10.1002/ese3.1304. Accessed 27 Sept. 2022.
Zhang, Fang, and Nuan Wen. "Carbon Price Forecasting: A Novel Deep Learning Approach." Environmental Science and Pollution Research, vol. 29, no. 36, 19 Mar. 2022, pp. 54782–54795, https://doi.org/10.1007/s11356-022-19713-x.
Zhang, Xiaobo, et al. "Short-Term Electric Load Forecasting Based on Singular Spectrum Analysis and Support Vector Machine Optimized by Cuckoo Search Algorithm." Electric Power Systems Research, vol. 146, May 2017, pp. 270–285, https://doi.org/10.1016/j.epsr.2017.01.035. Accessed 26 Mar. 2020.
Zhang, Xingmin, et al. "Carbon Trading and COVID-19: A Hybrid Machine Learning Approach for International Carbon Price Forecasting." Annals of Operations Research, 25 Apr. 2023, https://doi.org/10.1007/s10479-023-05327-0.
Zhu, Bangzhu, and Yiming Wei. "Carbon Price Forecasting with a Novel Hybrid ARIMA and Least Squares Support Vector Machines Methodology." Omega, vol. 41, no. 3, June 2013, pp. 517–524, https://doi.org/10.1016/j.omega.2012.06.005.

"Allowance Allocation | California Air Resources Board." Ww2.Arb.ca.gov, ww2.arb.ca.gov/our-work/programs/cap-and-trade-program/allowance-allocation.
"Allowance Price Explorer." International Carbon Action Partnership, icapcarbonaction.com/en/ets-prices. Accessed 18 Mar. 2024.
"Analysts Lowered EU Carbon Price Forecasts due to Weak Power Sector Demand | NEWS." Reccessary, www.reccessary.com/en/news/world-finance/analysts-lowered-eu-carbon-price-forecasts-due-to-weak-power-sector-demand. Accessed 27 Mar. 2024
"Annex III: Scenarios and Modelling Methods." 17 Aug. 2023, pp. 1841–1908, https://doi.org/10.1017/9781009157926.022.
"An Ensemble Prediction System Based on Artificial Neural Networks and Deep Learning Methods for Deterministic and Probabilistic Carbon Price Forecasting." Frontiers in Environmental Science, 17 Sept. 2021, www.researcher-app.com/paper/8762007. Accessed 27 Mar. 2024.
"Bloomberg - BCAX Index" Www.bloomberg.com, www.bloomberg.com/quote/BCAX:IND/members and "Bloomberg - BLAX Index" Www.bloomberg.com, www.bloomberg.com/quote/BLAX:IND/members. Accessed 19 Mar. 2024.
"Bloomberg - Bloomberg Increases Carbon Emissions Data to Cover 100,000 Companies" Www.bloomberg.com, www.bloomberg.com/company/press/bloomberg-increases-carbon-emissions-data-to-cover-100000-companies/. Accessed 18 May 2024.
"Carbon." ICIS Explore, www.icis.com/explore/commodities/energy/carbon/. Accessed 8 Apr. 2024.
"Carbon Market Data - Home." Www.carbonmarketdata.com, www.carbonmarketdata.com/en/home. Accessed 8 Apr. 2024.
"Carbon Markets | S&P Global." Www.spglobal.com, www.spglobal.com/esg/solutions/carbon-markets.
"China - Beijing Pilot ETS | International Carbon Action Partnership." Icapcarbonaction.com, icapcarbonaction.com/en/ets/china-beijing-pilot-ets.
"China -Shanghai Pilot ETS General Information." https://icapcarbonaction.com/system/files/ets_pdfs/icap-etsmap-factsheet-62.pdf
"CICC HKAM." Cicchkam.com, cicchkam.com/en/product/documents/4?pro=9. Accessed 19 Mar. 2024.
"Climate Change AI - About." Climate Change AI, www.climatechange.ai/about#people.
"Compare ETS | International Carbon Action Partnership." Icapcarbonaction.com, icapcarbonaction.com/en/compare/53/62. Accessed 27 Mar. 2024.
"Eikon." Eikon.refinitiv.com, eikon.refinitiv.com/.
"Elements of RGGI | RGGI, Inc." Rggi.org, 2023, www.rggi.org/program-overview-and-design/elements.
"European Energy Exchange AG (EEX)." Www.eex.com, www.eex.com/en/.
"European Environment Information and Observation Network." Eionet Portal, www.eionet.europa.eu/.
"European Industrial Emissions Portal." Industry.eea.europa.eu, industry.eea.europa.eu/.
"HANetf." HANetf, www.hanetf.com/product/30/fund/sparkchange-physical-carbon-eua-etc. Accessed 19 Mar. 2024.
"Harbor Capital | ETF | Harbor Energy Transition Strategy ETF (RENW)." Www.harborcapital.com, www.harborcapital.com/etf/renw/. Accessed 19 Mar. 2024.
"Homepage | California Air Resources Board." Ca.gov, 7 Nov. 2019, ww2.arb.ca.gov/.
"ICE Carbon Futures Index Family." Www.ice.com, www.ice.com/fixed-income-data-services/index-solutions/commodity-indices/carbon-futures. Accessed 8 Apr. 2024.
"IPath® Series B Carbon ETN, GRN:PCQ:USD Summary - FT.com." Markets.ft.com, markets.ft.com/data/etfs/tearsheet/summary?s=GRN:PCQ:USD. Accessed 19 Mar. 2024.
"IPCC. Mitigation of Climate Change Climate Change 2022 Working Group III Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. 2022. " https://www.ipcc.ch/report/ar6/wg3/downloads/report/IPCC_AR6_WGIII_FullReport.pdf
"Korea Emissions Trading Scheme General Information ETS Description." https://icapcarbonaction.com/system/files/ets_pdfs/icap-etsmap-factsheet-47.pdf
"Lawrence Goulder's Profile | Stanford Profiles." Profiles.stanford.edu, profiles.stanford.edu/lawrence-goulder?tab=publications. Accessed 27 Mar. 2024.
"NGFS." Banque de France, NGFS.net. Accessed 18 May 2024. http://NGFS.net
"NGFS Climate Scenarios Technical Documentation. 2023." https://www.ngfs.net/sites/default/files/media/2024/01/16/ngfs_scenarios_technical_documentation_phase_iv_2023.pdf
"NGFS Scenarios Portal." NGFS Scenarios Portal, www.ngfs.net/ngfs-scenarios-portal/.

"Neural Nets Learn to Program Neural Nets with Fast Weights (1991)." Idsia.ch, 2021, people.idsia.ch/~juergen/fast-weight-programmer-1991-transformer.html. Accessed 20 May 2024.
"Open Knowledge Repository." Openknowledge.worldbank.org, openknowledge.worldbank.org/collections/d9ab9a14-0168-54e9-9ff2-b84a1c83fb43?spc.sf=dc.date.issued&spc.sd=DESC. Accessed 19 Mar. 2024.
"Policies, Projections, and the Social Cost of Carbon: Results from the DICE- 2023 Model." Yale Department of Economics, economics.yale.edu/research/policies-projections-and-social-cost-carbon-results-dice-2023-model. Accessed 27 Mar. 2024.
"PRICES - AlliedOffsets." Alliedoffsets.com, 26 July 2022, alliedoffsets.com/prices-2/.
"Quantum Commodity Intelligence | Carbon News, Prices & Analysis." Www.qcintel.com, www.qcintel.com/carbon/. Accessed 8 Apr. 2024.
"Scott Barrett | Columbia SIPA." Www.sipa.columbia.edu, www.sipa.columbia.edu/communities-connections/faculty/scott-barrett. Accessed 27 Mar. 2024.
"South Korea to Launch Asia's First Retail Carbon-Linked Investment Product" www.bloomberg.com, www.bloomberg.com/news/articles/2024-03-07/south-korea-to-launch-asia-s-first-retail-carbon-linked-investment-product. Accessed 19 Mar. 2024.
"Tanpaifang.com - Shanghai ETS Information." http://www.tanpaifang.com/tanjiaoyi/2013/0712/22185.htm
"Tanpaifang.com - Shanghai ETS Information." https://www-tanpaifang-com/tanjiaoyi/2022/0922/90734.html
"_上海市生态环境局." Sthj.sh.gov.cn, sthj.sh.gov.cn/hbzhywpt2022/index.html. Accessed 18 May 2024.
"The Beginner's Guide to Representative Concentration Pathways." Skeptical Science, skepticalscience.com/rcp.php?t=3#concentrations. Accessed 18 May 2024.
"The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 2018." NobelPrize.org, www.nobelprize.org/prizes/economic-sciences/2018/nordhaus/facts/.
"Understanding GEC Model Scenarios – Global Energy and Climate Model – Analysis." IEA, www.iea.org/reports/global-energy-and-climate-model/understanding-gec-model-scenarios.
"USA -California Cap-And-Trade Program General Information ETS Description." https://icapcarbonaction.com/system/files/ets_pdfs/icap-etsmap-factsheet-45.pdf
"Voluntary Registry Offsets Database | Berkeley Carbon Trading Project | CEPP Projects | Center for Environmental Public Policy (CEPP) | Centers | Research and Impact | Goldman School of Public Policy | University of California, Berkeley." Gspp.berkeley.edu, gspp.berkeley.edu/research-and-impact/centers/cepp/projects/berkeley-carbon-trading-project/offsets-database.
"Welcome to Fred, Federal Reserve Economic Data. Your Trusted Source for Economic Data since 1991." FRED, Federal Reserve Bank of St. Louis, fred.stlouisfed.org/. Accessed 18 Mar. 2024.
"Welcome to the ICAP ETS Map." Icapcarbonaction.com, icapcarbonaction.com/en/ets.
"Welcome | RGGI, Inc." Rggi.org, 9 July 2019, www.rggi.org/.
"What Are Climate Model Phases and Scenarios? | USDA Climate Hubs." www.climatehubs.usda.gov, www.climatehubs.usda.gov/hubs/northwest/topic/what-are-climate-model-phases-and-scenarios.
"What Is a Pigouvian Tax?" Tax Foundation, 25 July 2023, taxfoundation.org/taxedu/glossary/pigouvian-tax/#:~:text=A%20Pigouvian%20tax%2C%20named%20after
"Why a Carbon Tax Alone Isn't Enough to Beat Climate Change." Nobel Perspectives, www.ubs.com/microsites/nobel-perspectives/en/latest-economic-questions/environmental-economics/articles/carbon-tax-plus-new-energy-technology-to-beat-climate-change.html. Accessed 27 Mar. 2024.
"Wolfram Schlenker." Columbia.edu, 2023, www.columbia.edu/~ws2162/research.html. Accessed 27 Mar. 2024.
"World Carbon Fund a | Carbon Cap Management LLP." RCM Alternatives, www.rcmalternatives.com/fund/world-carbon-fund-a-carbon-cap-management-llp/. Accessed 18 May 2024.
"World Carbon Pricing Database." Resources for the Future, www.rff.org/publications/data-tools/world-carbon-pricing-database/.
"Yoshua Bengio." Scholar.google.com, scholar.google.com/citations?user=kukA0LcAAAAJ&hl=en.
"北京市碳排放权电子交易平台." Bjets.com.cn, 2024, www.bjets.com.cn/article/jyxx/?2. Accessed 18 May 2024.
"北京市碳排放权电子交易平台." Bjets.com.cn, 2024, www.bjets.com.cn/article/xxzx/jysgg/. Accessed 18 May 2024.
"《上海市碳排放交易管理办法》_碳排放交易网——全球领先的碳市场门户中文网站." Www.tanpaifang.com, www.tanpaifang.com/tanjiaoyi/2013/0712/22185.html. Accessed 18 May 2024.