**Research Proposal: Explaining EU GO Prices**

**1  Purpose & Scope**

Provide a transparent, data‑driven framework to explain historical price movements in EU GO certificates (hydro, biomass, solar, wind) across spot and forward vintages (possible extension: to create scenario‑ready forecasts that inform trading, risk and structuring decisions).

**2  Core Questions & Hypotheses**

1. What fundamental and market variables drive GO price levels and volatility?
2. Does a stable long‑run equilibrium exist between GOs and carbon or fuel prices?
3. Is the relative spread Hydro ∕ Biomass vs Solar ∕ Wind a liquidity barometer?

**H1 Liquidity‑Spread Hypothesis:** A widening Hydro/Biomass premium signals tightening liquidity; a narrowing spread signals surplus.

**3  Data Architecture**

| **Layer** | **Dataset** | **Frequency** |  |
| --- | --- | --- | --- |
| **Prices** | Spot & forwards | Weekly |  |
| **Market Drivers** | TTF day‑ahead & forward strip, EUA Dec futures, Nordic/DE power baseload | Daily |  |
| **Production Fundamentals** | Installed capacity by tech, hydro reservoir indices (Nordic, Alpine), biomass pellet index, solar irradiance, wind‑speed anomalies, analyst forecasts of renewables share | Weekly–Monthly |  |
| **Microstructure** | Bid‑ask spread, trade volume, issue/cancel stats, banking rules, compliance deadlines | Daily–Monthly |  |
| **Compliance calendar** | Policy news0 | Flags |  |

**4  Methodological Framework**

**4.0 Setting up**

* Work with log differences of mid prices
* 1% change in regressor -> x% change in price

**4.1 Exploratory Analysis**

* Seasonality (monthly, weekly, daily level), correlations, summary statistics, test for trend/normality
* PCA on the forward curve to reduce multicollinearity (will check with Marc as he has worked on forward curves
* Break tests to identify where to split the data

**4.2 Econometric Layer**

* **ARIMAX–GARCH‑X** per tech with; exogenous variance terms: ΔEUA, ΔTTF, reservoir etc.
* **Markov‑Switching GARCH** (low/high‑vol). Dummy or regime probability flags crisis periods (2022‑23).
* **GARCH‑MIDAS (weekly × monthly)**: the short‑run component uses weekly squared returns; the long‑run MIDAS filter uses monthly reservoirs & macro variables.

**4.3 Machine‑Learning Layer (Benchmark & Non‑Linearity Check) (POSSIBLE EXTENSION)**

* Gradient Boosting (XGBoost) and compact LSTM, fed with engineered lags & interaction features.
* SHAP analysis to quantify variable importance and uncover thresholds (e.g. hydro storage < 30 %).

**4.4 Stress & Scenario Engine**

Might use Monte Carlo to simulate scenarios and test the model

**5  Regression Designs**

1) Baseline

* X\_t = {Δ EUA, Δ TTF, hydro‑reservoir z‑score, days‑to‑year‑end, …}.
* ε\_t follows a GARCH(1,1) process to capture conditional volatility.

2) Multi-tech panel

α\_i captures tech‑specific constants; δ\_t common time shocks.

3)SUR

This captures cross-tech correlation without forcing identical coefficients.

4) Extension: Forward-curve factor model (vintages 2024–2028)

Step 1 – factor extraction

At each date t, run PCA on the log-prices across vintages to obtain level, slope, curvature

Step 2 – factor dynamics

Optionally estimate as a VAR

| **Phase** | **Scope** | **Regression used** |
| --- | --- | --- |
| **1** | Hydro spot | Eq. 1 |
| **2** | All techs, spot | Eq. 2 (panel FE) / Eq. 3 (SUR) |
| 3 | Add forward vintages | Eq. 4 (factor model) |

‘