1 Data rationale

Table 1: Cocoa data: rationale and break-detection strategy

Component	**Description**
Data type and object	Cocoa is a traded commodity; we work with log-returns $Y_t = \Delta \log P_t$ and covariates X_t (e.g. weather variables). Following CGS, the series is viewed as a nonstationary process generated by a time-varying conditional mean function $m_t(x)$.
Structural breaks	In the CGS framework, nonstationarity arises through time-variation in $m_a(\cdot) \neq m_b(\cdot)$. Cocoa prices are plausibly nonstationary and subject to structural breaks driven by climate and supply-chain shocks (e.g. El nino supply disruptions), which induce discrete changes visually in the conditional mean/variance.
Economic interpretation of breaks	El nino is widely believed to be dominant the break in the conditional mean of log-returns. This aligns with the CGS notion of a structural break as a change in the regression function, and yields an economically interpretability before and after break
Break number and size	There are several identifiable breaks in the Cocoa log-return series, however, dominant by a single large break after 2024 El nino event which we argued earlier. This break has significant volatility and mean shifts. For sake of comparison, we will only validate the CGS framework until end of 2024.
Break detection strategy (Mohr and Selk, 2020)	To determine the break date τ in a data-driven way, Employ as Cai et al suggested, Mohr and Selk (2020). This is plausible under our assumption of Log return being strong mixing. Matches the break-detection step recommended in the CGS framework.
Within-regime approximation	Conditional on the estimated break date τ, we treat the process as piecewise stationary and strongly mixing within each regime(pre and post). This allows us to apply the CGS weighted local linear procedures and asymptotic theory locally within each regime.

In summary, the Cocoa series fits naturally into the CGS framework of nonparametric time-varying regression with a single structural break. A well designed experimental setup with baseline methods will allow us to validate the CGS method's effectiveness in this real-world scenario.

2 Experiment setup and models

Table 2: Proposed Model Architecture for Cocoa Price Forecasting

Group	Model Name	Data Usage	Mechanism & Rationale
1. NP Benchmarks	**Post-Break LL** (PBLL)	Post-break	**Low bias, high variance baseline** Uses Local Linear estimation only on new regime data. Low bias, potentially high variance.
	Pre-Break LL (PrBLL)	Pre-break	**High Bias, low var baseline** Assumes the old regime persists. Used to quantify the magnitude of the structural break ($\lambda(x)$).
	Weighted LL (WLL)	Combined	**Cai et al., 2025** Convex combination of Pre/Post estimators. Uses **MFV** to select γ, balancing bias from old data against variance of new data.
2. ML Competitors	**XGBoost**	Full Sample (with Time Index)	**Gradient Boosting Standard.** Tests if a "black box" learner can implicitly adapt to the structural break without explicit weighting.
	Random Forest	Full Sample (with Time Index)	**Bagging Standard.** Used to evaluate if lag is more pronounced in bagging methods compared to WLL.
3. ML Extension	**Weighted ML Ensemble**	Combined	**The Methodological Mirror.** Replaces Local Linear regressors with ML models. Explicitly applies the Cai et al. (2025) weighting strategy to ML outputs via MFV.

Combined means using Cai et al's linear combination between regimes idea.$\gamma \text{Pre} + (1-\gamma)\text{Post}$