3 questions I have

1. Looking at Equation (2.7), the WLL estimator $\hat{\beta}(x)$ acts as a convex combination of the pre-break and post-break estimators. My interpretation is that while the Local Linear estimator itself is first-order (fitting slopes), the mixing strategy via the weight matrix GAMMA is effectively 'zero-order'—it calculates a scalar weighted average of the two regime parameters.

Does this explain why the bias term involving the break size, $s_b\lambda(x)$, could not be diminished, as structural break setting, $\gamma > 0$? It seems the estimator forces a "compromise line' between the two regimes, and the only way the method solves this is by using the MFV to detect when this is just too costly (large λ), which will force $\gamma \to 0$?"

2. Regarding the high-dimensional limitation ($d < 5$) you mentioned in the conclusion: is it accurate to interpret this 'curse of dimensionality' as the local window simply failing to capture the full geometry of $m_t(\cdot)$? i.e using finite amount of parameters to capture a infinite diemensional non-parametric function. Does this collapsion of infiite-NP into finite-P setting make both NP and P weaker than ML methods? This is relevant because I plan to extend your framework for my Cocoa project using a Machine Learning model (like Random Forest) to handle high-dimensional indicators. I am wondering: if I replace the local linear estimator with a 'black box' ML model, would the asymptotic optimality of the Multifold Forward-Validation weights (Theorem 2.2) likely be preserved, or would that require a completely new consistency proof?

3. The paper's conclusion notes that constructing Prediction Intervals is a direction for future work. Given that my current Cocoa Project aims to quickly establish a benchmark comparison within a two-month deadline, I would like to ask for your advice: If I temporarily ignore variance breaks and interval forecasting to focus solely on comparing point forecast MSFE, would this simplified setup be sufficient for an 'entry-level project'? Or, in your opinion, is a mean forecast comparison not particularly meaningful for financial data without also modeling risk and volatility?

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** (with Time Index)	Full Sample	**Gradient Boosting Standard.** Tests if a "black box" learner can implicitly adapt to the structural break without explicit weighting.
	Random Forest (with Time Index)	Full Sample	**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. $\gamma \text{Pre} + (1 - \gamma)\text{Post}$

2.1 Data preprocess

Drop daily highest/lowest temperature records, there are no further missing values in Precipitation and Average temperature. Then since more than 1 station might record data for the same day, we will use anomoly valuse averaging from past data to aggregate them into a single daily record without peaking ahead.

After that we create log price and log return from daily price data. Then we merge weather data with log-return data by date and drop the missing values.

2.2 Model strucuture

We strictly follow the CGS framework to set up our models. But aware by the point out of Cai et al, that the recommended dimension will be less than 5. So this strictly limits the number of past lags we can use as features. We will also strictly follow the CGS cross-validation procedure to select hyperparameters. i.e MFV set to $Q = 4, etc$.

3 Evaluation metrics

We use a primary MSFE as Cai et al 2025 did, focusing on OOS performance in the post-break regime: Secondary metric will be using MDM to compare the performance between models pairwise.

3.1 Hypothesis

While ML models without break weighting strategy may achieve low overall MSFE, however, WLL estimator will demonstrate fast responding adatpion in the immediate post-break period by lower cumulative error. This would suggest the explicit weighting strategy is effective in handling structural breaks in nonparametric time series forecasting.

i.e penalize $\lambda(x)$ quick enough to diminish the bias term if necessary.