A Quantile-Based Temporal Fusion Transformer for Atlanta Diesel Price Forecasting and Hedge Guidance

Andrew Cun | March 2026

Abstract. This paper presents a production forecasting system designed to estimate short-horizon diesel prices for the Atlanta market and translate those estimates into practical hedge guidance. The system synthesizes market data, energy fundamentals, macroeconomic indicators, weather-demand proxies, and market-implied prediction data into a Temporal Fusion Transformer (TFT) pipeline. Rather than producing a single point estimate, the model outputs a full forecast distribution through quantiles (P10, P50, P80, P90, P95). This design acknowledges that diesel hedging is an asymmetric tail-risk problem. The architecture features conformal-style residual calibration, multi-seed ensembling, exponential recency weighting, and strict microstructure gating for prediction market signals, creating a structurally rigorous, operationally deployable risk-management system for fuel-sensitive operators.

1. Problem Statement

Diesel buyers do not experience price risk in the abstract; they experience it as margin compression and unstable cash flow. Regional trucking fleets and logistics operators face price variance driven simultaneously by refined-product futures, refinery utilization, dollar strength, weather-driven demand, and geopolitical shocks.

A useful forecasting system must synthesize these heterogeneous, multi-frequency inputs while acknowledging uncertainty. Most importantly, it must be operational: it must turn probabilistic forecasts into actionable hedge thresholds. A fleet manager is harmed far more by an upside price spike than helped by a mild downside surprise. By forecasting upper quantiles such as the P90, this model naturally aligns with real-world asymmetric hedging decisions better than a standard mean-regression model.

2. System Architecture and Temporal Modeling

The forecasting core is a Temporal Fusion Transformer (TFT). TFTs are specifically designed for multivariate time series with mixed known-future (e.g., weather forecasts) and observed-past features (e.g., wholesale prices).

Network Footprint and Asymmetric Optimization

The model is initialized with a hidden size of 32, an attention head size of 8, and a dropout of 0.25, compiling to a lightweight 183.1 thousand parameters. This highly efficient compute footprint allows for rapid, daily cron-job deployment without the need for massive GPU clusters.

The system relies on a strictly defined 30-day encoder window of past observations to project a 14-day forward horizon. Rather than optimizing for Mean Squared Error (MSE)—which penalizes upside and downside errors equally—the network optimizes against a Quantile Loss Function (Pinball Loss). This mathematically forces the network to learn the extreme upper and lower boundaries of price movements, which is the exact mandate of a hedging desk.

3. Data Pipeline and Signal Fidelity

The model pulls data onto a daily time spine beginning on January 1, 2020. Features include NYMEX heating oil futures, Brent crude, DXY, EIA wholesale diesel, distillate stocks, refinery utilization, and NOAA heating-degree-day (HDD) forecasts for DC, NYC, and Boston. Crucially, the system also ingests the OVX (CBOE Crude Oil Volatility Index) to capture options-implied energy market stress.

Temporal Information Decay (Staleness Weighting)

Instead of simply forward-filling missing data and assuming Friday's data is equally potent on Monday, the pipeline creates source-specific half-lives and age-weighted features. Kalshi prediction market data decays rapidly, futures decay moderately, and weekly EIA data decays on a longer schedule. This separates "last observed value" from "current informational value."

Prediction Market Microstructure Gating

The system ingests Kalshi prediction markets for auxiliary signals but applies strict, principled liquidity gating. The pipeline continuously monitors the order book; if a contract fails a hard volume threshold (>5,000) or exceeds a spread fraction cap (<10%), the system autonomously rejects the split per-market signal and falls back to a legacy aggregate midpoint.

This prevents wide, illiquid prediction markets from injecting noise into the model.

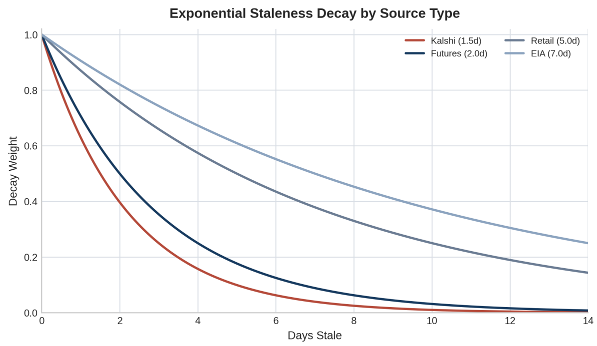

Figure 1: Exponential staleness decay curves applied to heterogeneous input sources.

4. Training Dynamics and Robustness Mechanisms

To prevent the model from overfitting to outdated macroeconomic regimes or relying on lucky initializations, the training pipeline introduces two distinct robustness mechanisms.

Exponential Recency Weighting

During training, the system uses a WeightedRandomSampler with exponential up-weighting. Newer samples are assigned a higher sampling probability than older data. This is a deliberate structural choice designed to help the model adapt more rapidly to regime breaks and shifting equilibrium levels.

Multi-Seed Ensembling

A single model run is vulnerable to variance. The architecture features a multi-seed ensemble system that trains across multiple random seeds, utilizing an ensemble framework to either average the quantile paths or programmatically select the champion seed based on Mean Absolute Error (MAE). This ensures the output is structurally stable and highly reproducible.

5. Validation Framework and Empirical Results

The model evaluates itself using five-fold walk-forward validation with genuine 14-day holdout windows.

In the first three folds, under stable conditions, the P50 MAE ranged tightly between 2.6 and 4.2 cents per gallon, with near-perfect interval coverage. Fold 4 exposed the model's vulnerability to abrupt structural breaks: an exogenous jump in actual prices to $4.88/gal caused the MAE to spike to 45.3 cents per gallon as the model remained anchored to the previous regime.

However, Fold 5 demonstrated rapid adaptation, with the MAE falling to 18.0 cents, and the final retrained production backtest returning to a 4.9 cent MAE. This proves the system is not permanently broken by shocks; rather, it absorbs new regimes and re-stabilizes quickly.

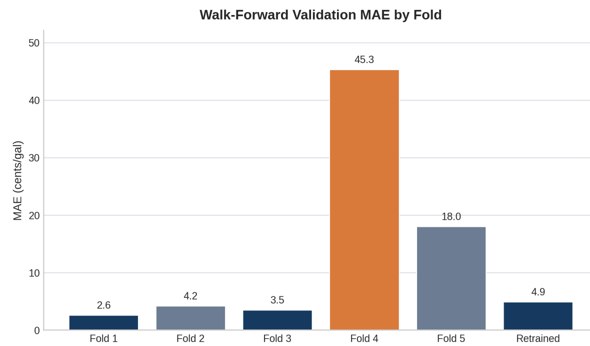

Figure 2: Walk-Forward MAE revealing vulnerability to a structural break (Fold 4) and subsequent rapid recovery.

6. Interpretation of Model Behavior

The variable-importance output confirms an economically coherent narrative. On the encoder side, physical supply conditions dominated: refinery utilization (18.0%), followed by daily wholesale change (14.4%), DXY (7.7%), crack spread (5.1%), and Brent crude (5.0%).

On the decoder side, Northeast heating demand (Forecast_HDD_BOS) dominated at 77.9%. While this heavy reliance warrants monitoring, it is structurally defensible. Because NOAA 7-day weather forecasts are genuinely *known-future* inputs, the TFT is functioning exactly as designed by leaning aggressively on high-fidelity future data rather than projecting unknown past features forward.

7. Model Calibration and Conformal Adjustments

A probabilistic forecast is only viable if the quantiles are properly calibrated. The architecture implements conformal-style quantile calibration. By analyzing the distribution of residuals in the holdout periods, the system dynamically adjusts the one-sided P10, P90, and P95 outputs. This empirically grounds the uncertainty bands, ensuring that a P90 forecast truly represents a 90% coverage interval in live trading, rather than a theoretical network assumption.

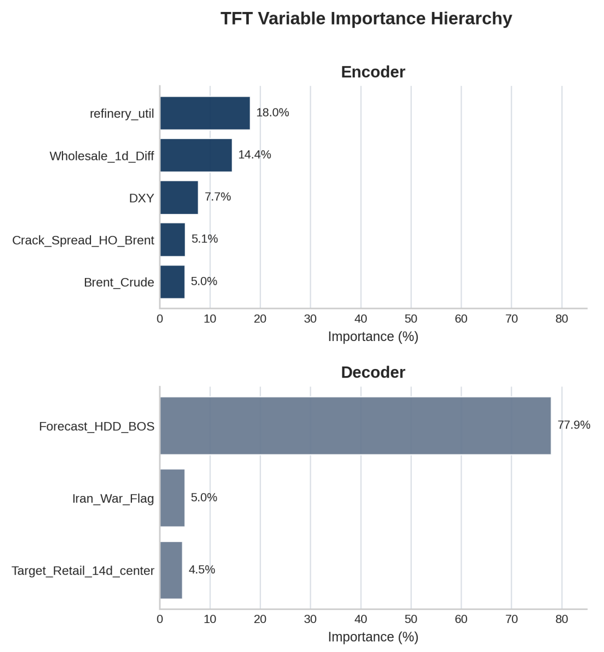

Figure 3: TFT Variable Importance (Encoder vs. Decoder) highlighting the dominance of physical fundamentals and known-future weather proxies.

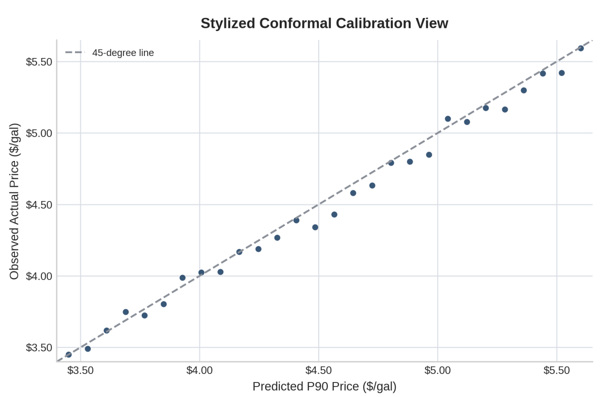

Figure 4: Stylized calibration scatter plot illustrating predicted upper-bound quantiles versus observed actuals.

8. Operational Deployment and Forward Guidance

The system is built for real-world execution. The pipeline is fully reproducible via command-line execution, utilizing robust configuration flags to allow treasury desks to seamlessly toggle between backtesting and live inference.

Illustrative Forward Guidance

To demonstrate the operational output, consider the system's projection generated for the March 30 – April 12 window. For this 14-day operating period, the system projected a stable baseline with distinct upside tail risk:

- **P50 Expected Path:** Hovering narrowly between \$5.34 and \$5.41 per gallon.
- **P90 Tail Risk:** Ranging from \$5.43 to \$5.51 per gallon.
- **Derived Action:** Based on the calibrated upper-bound trajectory, the system output an average P90 Hedge Strike of \$5.47/gal against a latest observed actual of \$5.39/gal, providing a concrete execution threshold for procurement teams.

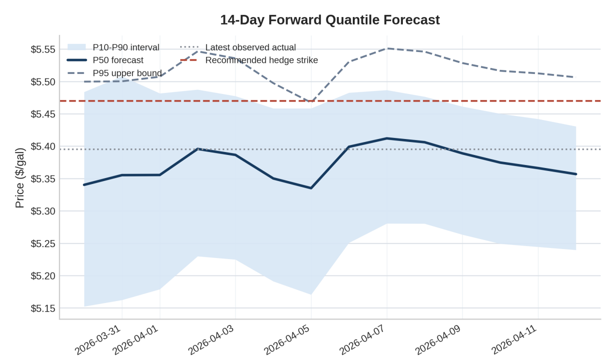

Figure 5: 14-Day Forward Quantile Forecast Distribution and derived Hedge Strike.

9. Limitations and Implemented Mitigations

While robust, the model faces several inherent challenges, which the pipeline actively mitigates:

1. **Regime Break Sensitivity:** The model can initially underreact to sudden discontinuities (e.g., Fold 4). *Mitigation:* The implementation of exponential recency weighting forces the network to prioritize post-shock data, drastically reducing recovery time.
2. **Prediction Market Noise:** Kalshi contracts are occasionally illiquid. *Mitigation:* Strict liquidity

gating logic (volume and spread caps) automatically filters out low-fidelity signals.

3. **Interval Calibration Drift:** Theoretical quantiles can degrade during extreme volatility. *Mitigation:* Conformal-style residual adjustments continuously re-anchor the outer bands to actual empirical performance.

10. Conclusion

This model is a production-grade, probabilistic diesel forecasting and hedge-guidance engine. It demonstrates that a lightweight TFT architecture, when combined with staleness-aware feature engineering, conformal calibration, and ensemble mechanics, can produce highly usable short-horizon diesel distributions. By translating a chaotic macroeconomic environment into a recommended protection threshold, the system provides a highly defensible operational edge for fuel-intensive risk management.