Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

RESEARCH ARTICLE

Bayesian Parameter Uncertainty in WTI Average Price Option Valuation

Heriberto Espino Montelongo

[1]Universidad de las Américas Puebla, Puebla, Mexico

Correspondence
Corresponding author: Heriberto Espino Montelongo.
Email: [corresponding email]

Abstract

Parameter uncertainty is often treated as secondary in derivative valuation, even though volatility is estimated from finite and potentially unstable samples. This paper develops a Bayesian framework that separates statistical learning under the physical measure from no-arbitrage valuation under the risk-neutral measure and propagates posterior uncertainty in volatility into arithmetic-average option prices. The framework is evaluated first in repeated-sampling experiments, where the true risk-neutral price is known, and then in an empirical application to CME WTI Average Price Options. The empirical database contains 213 unique call and put contracts and 11,179 option-date observations spanning expiries from September 2026 to June 2029. The contract-specific implementation respects the calendar-month average of first-nearby WTI futures settlements and distinguishes realized from remaining fixings. Preliminary synthetic evidence indicates that posterior integration and posterior-mean plug-in pricing can be nearly indistinguishable in regular settings, while MAP pricing is less stable in small samples. The empirical design asks when that apparent irrelevance breaks down as maturity, moneyness, posterior dispersion, and oil-market volatility change.

KEYWORDS

Bayesian option pricing; Asian options; WTI crude oil; parameter uncertainty; commodity derivatives

Internal working draft — not for circulation. Empirical pricing results remain in progress.

1 | INTRODUCTION

Derivative values are conditional on parameters that are not observed. In the Black–Scholes–Merton benchmark, volatility is commonly inserted into the pricing map as though it were known, even when it has been estimated from a finite sample of returns. This plug-in convention is operationally convenient, but it suppresses a distinct source of uncertainty: uncertainty about the parameter that enters the risk-neutral valuation rule. A Bayesian treatment makes this distinction explicit by replacing a single volatility estimate with a posterior distribution and propagating that distribution through the derivative-pricing function.

The distinction is especially relevant for path-dependent commodity derivatives. Average-price options depend on a sequence of underlying prices rather than a terminal observation, and their sensitivity to volatility changes with moneyness, maturity, and the fraction of the averaging window that has already been fixed. Commodity markets add a further complication because the economically relevant state is a futures curve rather than a single spot asset. These features create a natural setting in which posterior parameter uncertainty may be economically negligible in some regions of the contract space but material in others.

This paper studies that question using CME WTI Average Price Options (APOs). The contract pays against the arithmetic average of daily first-nearby WTI futures settlements over a calendar month. Consequently, a faithful valuation exercise must account for both the first-nearby roll convention and partial fixing: once the averaging month has begun, some components of the final average are known while the remaining components are stochastic. Treating the option as an ordinary European claim on a single fixed-maturity futures contract would therefore change the object being valued.

The central methodological principle is a strict separation between inference under the physical measure and valuation under the risk-neutral measure. Historical returns are used to infer the parameters of a data-generating process under $\mathbb{P}$. The resulting posterior uncertainty in volatility is then mapped into option values under $\mathbb{Q}$. The physical drift is not inserted into the risk-neutral pricing dynamics. This distinction is elementary in no-arbitrage theory Black and Scholes (1973), Merton (1973), but it becomes easy to obscure when estimation and Monte Carlo valuation are implemented in a single numerical pipeline.

Let $C^{\mathbb{Q}}(\sigma)$ denote the risk-neutral value of an arithmetic-average call conditional on volatility σ. The full-Bayes valuation rule is

$$C_{FB} = \mathbb{E}\left[C^{\mathbb{Q}}(\sigma) \mid \mathcal{D}\right], \tag{1}$$

whereas the posterior-mean plug-in rule is

$$C_{PM} = C^{\mathbb{Q}}\left(\mathbb{E}[\sigma \mid \mathcal{D}]\right). \tag{2}$$

The two need not coincide because the option-pricing map is nonlinear. A second-order expansion around $\bar{\sigma} = \mathbb{E}[\sigma \mid \mathcal{D}]$ gives the approximation

$$C_{FB} - C_{PM} \approx \frac{1}{2} C^{\mathbb{Q}\prime\prime}(\bar{\sigma}) \operatorname{Var}(\sigma \mid \mathcal{D}), \tag{3}$$

which isolates two forces: posterior dispersion and curvature of the pricing function. Equation (3) motivates the empirical analysis. Parameter uncertainty should matter most where the posterior is diffuse and where option value is highly curved in volatility.

The paper uses two complementary forms of evidence. First, repeated-sampling experiments provide a controlled environment in which the true volatility and the associated risk-neutral option value are known. This makes it possible to evaluate bias, mean absolute error, root mean squared error, and posterior interval coverage without defining the target from the posterior itself. That distinction avoids the circular comparison that arises when an estimator is evaluated against draws generated from the same posterior. Second, the paper develops an empirical application using historical CME WTI APO settlements obtained through Barchart. The current panel contains 213 unique call and put contracts and 11,179 option-date observations across expiries from September 2026 through June 2029.

The empirical design deliberately uses the full downloaded contract panel rather than selecting a handful of strikes according to current moneyness. This is important in the recent WTI sample because the level of crude oil changed sharply over the life of many contracts. A strike that is out of the money on the final observation date may have been near or in the money earlier in its history. Moneyness is therefore constructed at the option-date level using the expected final average, not assigned permanently from one market snapshot. Liquidity variables are retained as filters and robustness dimensions rather than used as the sole criterion for selecting contracts.

The paper contributes to three related literatures. The first is Bayesian option pricing and parameter uncertainty. Existing work shows how posterior parameter and state uncertainty can be incorporated into predictive and risk-neutral densities Guidolin and Timmermann (2003), Rombouts and Stentoft (2014). The second is the valuation of Asian and average-price options, where arithmetic averaging generally eliminates the simple closed form available for standard European options and motivates conditioning, approximation, and simulation methods Kemna and Vorst (1990), Curran (1994), Kahalé (2020). The third is commodity-derivative valuation, where contract mechanics and the term structure of futures prices are central to the economically relevant pricing state. The present paper connects these strands by asking a narrower empirical question: when does integrating parameter uncertainty improve the valuation of a traded path-dependent energy derivative relative to conventional plug-in rules?

An important feature of the analysis is that it allows a negative result. Preliminary synthetic evidence indicates that full posterior integration and the posterior-mean plug-in can be almost identical in regular Black–Scholes settings. That finding is not treated as a failure of the Bayesian framework. Instead, it establishes a benchmark: if the pricing function is locally close to linear in the posterior support and the posterior of volatility is concentrated, integration should add little. The WTI application is then designed to identify the contract and market states in which this approximation ceases to be adequate.

The remainder of the paper is organized as follows. Section 2 describes the WTI Average Price Option and the empirical panel. Section 3 presents physical-measure inference, risk-neutral valuation, and the competing pricing rules. Section 4 describes the synthetic and empirical research designs. Section 5 reports the currently available evidence and defines the remaining empirical tests. Section 6 sets out robustness checks and extensions, and Section 7 concludes.

2 | WTI AVERAGE PRICE OPTIONS AND THE EMPIRICAL SETTING

2.1 | Contract mechanics

CME WTI Average Price Options are cash-settled options whose payoff is linked to the arithmetic average of daily settlements of first-nearby WTI crude-oil futures during the relevant calendar month. For an averaging month with fixing dates $t_1, \ldots, t_M$, define

$$A_T = \frac{1}{M} \sum_{j=1}^{M} F_{t_j}^{(1)}, \tag{4}$$

where $F_{t_j}^{(1)}$ denotes the settlement of the first-nearby CL futures contract applicable to fixing date t_j. A call and put have terminal payoffs

$$H_C = (A_T - K)^+, \tag{5}$$
$$H_P = (K - A_T)^+. \tag{6}$$

The first-nearby designation matters because the futures contract contributing to the average can change during the averaging period. The payoff is therefore not, in general, an average of repeated observations of one fixed futures maturity.

At an option valuation date t inside the averaging month, let $\mathcal{R}_t$ denote fixing dates already realized and $\mathcal{U}_t$ the remaining fixing dates. The final average can be decomposed as

$$A_T = \frac{1}{M} \left(\sum_{j \in \mathcal{R}_t} F_{t_j}^{(1)} + \sum_{j \in \mathcal{U}_t} F_{t_j}^{(1)} \right). \tag{7}$$

This decomposition is central to the empirical pricing exercise. Realized fixings are known constants; only the remaining components are simulated under the risk-neutral measure. As the averaging month progresses, the option mechanically loses exposure to future volatility because a larger share of A_T has already been fixed.

A useful state variable for cross-sectional analysis is the risk-neutral expected final average,

$$\widehat{A}_{t,T}^{Q} = \frac{1}{M} \left(\sum_{j \in \mathcal{R}_t} F_{t_j}^{(1)} + \sum_{j \in \mathcal{U}_t} \mathbb{E}_t^{Q}[F_{t_j}^{(1)}] \right). \tag{8}$$

We define log moneyness observation by observation as

$$m_{t,K,T} = \log \left(\frac{K}{\widehat{A}_{t,T}^{Q}} \right). \tag{9}$$

For calls, negative values correspond to strikes below the expected average; for puts the economic interpretation is reversed in terms of intrinsic direction. This measure is preferable to assigning a contract a fixed moneyness category using the futures price on the final sample date.

2.2 | Why WTI is informative for parameter uncertainty

WTI provides a demanding empirical environment for the question studied in this paper. Oil prices can exhibit rapid changes in level, realized volatility, and the shape of the futures curve. These changes imply that the same option can move through substantially different moneyness states during its observed history. They also create episodes in which a posterior estimated from a rolling historical window may be materially more dispersed than in quieter periods.

The analysis does not attribute individual price moves mechanically to particular geopolitical events. Such causal statements would require a separate event-study design. Instead, market turbulence is measured directly from observed WTI returns. Dates are classified into volatility regimes using rolling realized volatility, allowing the pricing comparison to ask whether posterior integration becomes more relevant when the data themselves indicate an unusually volatile market state.

2.3 | Barchart option histories

The empirical database was assembled from individual Barchart historical series for WTI Average Price Options. The downloaded maturities are September 2026, October 2026, November 2026, March 2027, September 2027, March 2028, September 2028, and June 2029. Calls and puts were retained as available. Rather than choosing only contracts that appear near the money on one date, the raw panel keeps all downloaded contracts and lets moneyness vary by date according to Equation (9).

The ingestion pipeline recursively audits the raw CSV directory, parses option symbols, derives strike and option type, checks the symbol-implied expiry against the storage folder, and de-duplicates contracts before constructing the option-date panel. The current audit identifies 214 input files, 213 unique contracts, and 11,179 option-date observations. One file is stored under a folder inconsistent with its symbol-implied expiry; the parser records the mismatch and uses the contract symbol rather than silently trusting the directory name.

The histories contain end-of-day fields including open, high, low, latest, volume, and open interest. A validation exercise found exact agreement between at least one Barchart `Latest` observation and the corresponding CME daily settlement for the same WTI APO contract and date. The empirical work therefore treats the field as an end-of-day settlement or mark after contract/date validation, not as evidence of a transaction on every date. This distinction is necessary because many histories contain zero reported volume while open interest remains positive.

Liquidity is consequently used as a diagnostic and robustness dimension rather than a binary requirement that every observation have positive trading volume. The main analysis will report results under alternative open-interest thresholds and will separately identify observations at the minimum quoted price. Contracts with sparse trading can still contain an exchange settlement, but conclusions that depend exclusively on such observations will be treated cautiously.

The data pipeline also produces a variable called `richest`, inherited from an earlier data-audit routine. It measures information completeness using the number of sessions, calendar span, and field completeness. It does *not* measure liquidity, trading intensity, or economic importance. The variable is retained for data-quality diagnostics but is not used as the economic contract-selection rule.

2.4 | Maturity coverage

The chosen expiries deliberately span the maturity dimension rather than concentrating exclusively on consecutive nearby months. The first three maturities provide short-horizon contracts, while March 2027 through June 2029 extend the cross section toward medium and long horizons. This structure is useful because Equation (3) predicts that the importance of posterior uncertainty depends not only on posterior dispersion but also on the curvature of the pricing map, which can change materially with time to maturity.

The number of available strikes and the depth of trading are not constant across maturities. In particular, the far end of the curve contains fewer contracts. The empirical design does not force a balanced panel. Instead, it preserves the observed contract set, reports coverage by maturity and option type, and uses weighting and sample restrictions as robustness exercises.

3 | METHODOLOGY

3.1 | Inference under the physical measure

The baseline statistical model assumes that the observed underlying follows a geometric Brownian motion under the physical measure,

$$dS_t = \mu S_t\, dt + \sigma S_t\, dW_t^P. \tag{10}$$

For equally spaced observations with interval Δt, log returns satisfy

$$R_i = \log\left(\frac{S_{t_i}}{S_{t_{i-1}}}\right) \sim \mathcal{N}\left[\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t, \sigma^2 \Delta t\right]. \tag{11}$$

The likelihood is the product of the conditional return densities. The simple diffusion is intentionally used as a transparent benchmark: the paper's primary object is the incremental effect of parameter uncertainty, not a claim that constant-volatility GBM is a complete model of crude-oil dynamics.

The baseline prior specification is

$$\mu \sim \mathcal{N}(0, 1), \tag{12}$$

$$\sigma \sim \text{InvGamma}(2, 0.1), \qquad \sigma > 0. \tag{13}$$

Prior sensitivity is evaluated separately. Sampling is performed in (μ, η) with $\eta = \log\sigma$ so that positivity is automatic. The transformed log posterior is

$$\log\pi(\mu, \eta \mid \mathcal{D}) = \log L(\mu, e^{\eta}; \mathcal{D}) + \log p(\mu) + \log p(e^{\eta}) + \eta + C, \tag{14}$$

where the final η is the Jacobian term.

A symmetric random-walk Metropolis proposal is used,

$$\theta' = \theta + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \Sigma_q), \tag{15}$$

with $\theta = (\mu, \eta)$. The acceptance probability is

$$\alpha(\theta, \theta') = \min\left\{1, \frac{\pi(\theta' \mid \mathcal{D})}{\pi(\theta \mid \mathcal{D})}\right\}, \tag{16}$$

following the Metropolis–Hastings construction Metropolis et al. (1953), Hastings (1970). Multiple-chain diagnostics, effective sample size, and prior sensitivity are included in the final empirical implementation rather than treating a single acceptance rate as sufficient evidence of convergence.

3.2 | Risk-neutral valuation

Historical inference and derivative valuation are kept conceptually and computationally separate. Under the Black–Scholes–Merton benchmark,

$$dS_t = (r - q)S_t\,dt + \sigma S_t\,dW_t^Q. \tag{17}$$

The historical drift μ in Equation (10) does not enter Equation (17). Posterior uncertainty about μ is therefore relevant for physical forecasting but is not propagated into the no-arbitrage price in the baseline model.

For a discretely monitored arithmetic-average call with fixing dates $t_1, \ldots, t_M$, conditional risk-neutral value is

$$C^Q(\sigma) = e^{-r\tau}\mathbb{E}_\sigma^Q\left[(A_T - K)^+ \mid \mathcal{F}_t\right], \tag{18}$$

where $\tau = T - t$. The posterior for σ induces a posterior distribution over theoretical prices through the push-forward

$$p(C \mid \mathcal{D}) = \int \delta_{C^Q(\sigma)}(C)p(\sigma \mid \mathcal{D})\,d\sigma. \tag{19}$$

This distribution quantifies parameter uncertainty in a model price. It is not a predictive distribution of portfolio losses and is not interpreted as VaR or expected shortfall.

3.3 | WTI futures representation

For the WTI APO application, the state variables are the first-nearby futures fixings in Equation (4). A parsimonious risk-neutral benchmark for a futures price is

$$dF_u = \sigma F_u\,dW_u^Q, \tag{20}$$

with zero drift under the pricing measure for the appropriately margined futures contract in the stylized setup. The implementation does not replace the entire APO with one fixed CL contract. For each remaining fixing date it identifies the first-nearby contract applicable to that date and initializes the simulated fixing from the corresponding point of the observed futures term structure.

For a valuation date inside the averaging month, simulation starts from Equation (7): realized fixings enter deterministically and only the remaining first-nearby settlements are stochastic. This distinction implies that two contracts with the same strike and time to final settlement can have different effective volatility exposure if their realized-average components differ.

The baseline single-factor specification is deliberately interpretable. It can be extended to maturity-specific volatilities, correlated futures-curve factors, stochastic volatility, or mean-reverting commodity models. Those alternatives are treated as robustness extensions because changing the underlying dynamics and changing the treatment of parameter uncertainty are separate questions.

3.4 | Monte Carlo valuation

Arithmetic-average options generally lack the simple closed form available for geometric averages. We therefore evaluate Equation (18) by Monte Carlo. In the standard Asian benchmark, antithetic sampling and a geometric-average control variate are used following the logic of Kemna and Vorst (1990). Let

$$Y = e^{-r\tau}(A_T - K)^+, \tag{21}$$

$$X = e^{-r\tau}(G_T - K)^+, \tag{22}$$

where G_T is the geometric average and $\mathbb{E}[X]$ is available analytically. The control-variate estimator is

$$Y_{cv} = Y - \beta\{X - \mathbb{E}[X]\}, \qquad \beta = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}. \tag{23}$$

Common random numbers are used when evaluating nearby volatility values for the same contract. This reduces simulation noise in the mapping $\sigma \mapsto C^Q(\sigma)$ and makes the full-Bayes/plugin comparison primarily a comparison of statistical decision rules rather than independent Monte Carlo errors.

3.5 | Competing valuation rules

Four estimators are compared:

$$\widehat{C}_{FB} = \mathbb{E}[C^Q(\sigma) \mid \mathcal{D}], \tag{24}$$

$$\widehat{C}_{PM} = C^Q(\mathbb{E}[\sigma \mid \mathcal{D}]), \tag{25}$$

$$\widehat{C}_{MAP} = C^Q(\widehat{\sigma}_{MAP}), \tag{26}$$

$$\widehat{C}_{MLE} = C^Q(\widehat{\sigma}_{MLE}). \tag{27}$$

The first estimator integrates the pricing map over the posterior. The second collapses the posterior to its mean before pricing. The third uses the posterior mode, and the fourth is a non-Bayesian likelihood benchmark.

A useful diagnostic is

$$\Delta_{FB,PM} = \widehat{C}_{FB} - \widehat{C}_{PM}. \tag{28}$$

A second-order expansion gives

$$\Delta_{FB,PM} \approx \frac{1}{2} C^{\mathbb{Q}\prime\prime}(\bar{\sigma}) \operatorname{Var}(\sigma \mid \mathcal{D}), \qquad \bar{\sigma} = \mathbb{E}[\sigma \mid \mathcal{D}]. \tag{29}$$

Equation (29) supplies the main mechanism tested in both the synthetic and empirical sections: posterior integration can matter only to the extent that posterior uncertainty interacts with nonlinear exposure to volatility.

4 | RESEARCH DESIGN

4.1 | Synthetic repeated-sampling experiment

The simulation study is designed to answer a question that cannot be answered cleanly with market prices alone: when the data-generating volatility is known, how much is lost by collapsing posterior uncertainty before pricing? Historical return samples are generated under $\mathbb{P}$ and option values are evaluated under $\mathbb{Q}$, preserving the measure separation in Section 3.

The baseline configuration fixes

$$\begin{aligned} S_0 &= 100, \quad \mu_0 = 0.08, \quad r = 0.03, \\ q &= 0, \quad \Delta t = 1/252. \end{aligned} \tag{30}$$

The production design varies the amount of historical information,

$$n \in \{63, 252, 1260\}, \tag{31}$$

true volatility,

$$\sigma_0 \in \{0.15, 0.25, 0.40\}, \tag{32}$$

moneyness, and option maturity. The contract grid is chosen to cover in-, near-, and out-of-the-money states over short, intermediate, and longer horizons. Posterior inference is performed once for each simulated history and then reused across contractual terms; changing a strike or maturity does not generate a new historical posterior.

For each scenario the external target is

$$C_0 = C^{\mathbb{Q}}(\sigma_0), \tag{33}$$

computed from a high-precision risk-neutral pricing grid. Importantly, C_0 is not a posterior draw and is not constructed from any of the estimators being compared. This prevents the mechanical advantage that the posterior mean would obtain if squared error were evaluated against the posterior distribution that defines it.

For method $M \in \{FB, PM, MAP, MLE\}$ and R independent histories, the main statistics are

$$\text{Bias}_M = R^{-1} \sum_{j=1}^{R} (\widehat{C}_{M,j} - C_0), \tag{34}$$

$$\text{MAE}_M = R^{-1} \sum_{j=1}^{R} |\widehat{C}_{M,j} - C_0|, \tag{35}$$

$$\text{RMSE}_M = \left[R^{-1} \sum_{j=1}^{R} (\widehat{C}_{M,j} - C_0)^2 \right]^{1/2}. \tag{36}$$

We additionally record coverage of central 95% posterior price intervals and analyze $\Delta_{FB,PM}$ directly. The production simulation uses deterministic seeds, restartable checkpoints, and shared pricing grids so that each reported result can be reproduced from the code version and experiment manifest.

4.2 | Empirical option panel

The raw empirical sample contains every downloaded WTI APO history rather than a hand-selected set of strikes. The unit of observation is option i on trading date t, identified by option type, strike, expiry month, and symbol. The raw panel currently contains 213 unique contracts and 11,179 option-date observations. The sample is intentionally unbalanced because strike availability and liquidity decline at longer maturities.

Three layers are kept separate. The *raw panel* preserves all parsed observations. The *main empirical panel* applies transparent data-quality and pricing-availability rules that can be reproduced from code. The *representative-contract set* is used only for figures and economic discussion. A contract can therefore be visually representative without determining the statistical sample.

The main sample does not require positive daily volume because an exchange settlement can exist on a day with no reported trade. Instead, volume, open interest, and minimum-tick indicators enter the diagnostics and robustness analysis. Alternative samples impose progressively stronger open-interest thresholds and remove observations pinned at the minimum quoted price. Results are reported with observation counts after every filter.

4.3 | Reconstructing first-nearby fixings and moneyness

Each option-date observation must be linked to the information set required by the actual APO payoff. The empirical pipeline therefore reconstructs: (i) realized first-nearby settlements for fixing dates already elapsed, (ii) the relevant CL futures contracts for remaining fixing dates, and (iii) the contemporaneous futures term structure used to initialize those remaining fixings. This mapping produces $\widehat{A}^{Q}_{t,T}$ in Equation (8) and the observation-specific moneyness measure in Equation (9).

This step is deliberately separated from the option-history ingestion. A generic continuous futures ticker is useful for exploratory historical-return analysis but is not treated as a contractual substitute for the sequence of first-nearby fixings that determines the APO payoff. The final empirical tables will report the source and roll convention for the CL curve used in valuation.

4.4 | Rolling parameter uncertainty

The empirical comparison requires a posterior available at each valuation date rather than a single posterior estimated using the full sample. Let $\mathcal{D}_t$ denote the historical information set available at date t. For a fixed or expanding estimation window, the posterior

$$p(\sigma \mid \mathcal{D}_t) \tag{37}$$

is estimated without using future returns. Each option-date observation is then priced by the four rules in Equation (27). This creates a panel of model prices and posterior diagnostics aligned with observed option settlements.

The principal out-of-sample pricing errors are

$$e^{M}_{i,t} = P^{M}_{i,t} - P^{mkt}_{i,t}, \tag{38}$$

where $P^{mkt}_{i,t}$ is the validated end-of-day settlement or mark and M indexes the pricing rule. Absolute and squared errors are aggregated overall and by economically relevant states. Because option settlements can themselves contain microstructure or administrative components, results are also stratified by open interest and trading activity.

4.5 | Volatility regimes and heterogeneous effects

Recent WTI history contains unusually large changes in price and volatility. Rather than label dates according to news narratives, the baseline regime variable is constructed from realized returns. For a 21-session window,

$$RV^{(21)}_t = \sqrt{252}\ \mathrm{sd}(r_{t-20}, \ldots, r_t). \tag{39}$$

Dates are classified into low-, medium-, and high-volatility regimes using sample quantiles determined without reference to option-pricing errors.

The mechanism in Equation (29) is examined by relating the full-Bayes/plugin gap to posterior variance, moneyness, maturity, partial-fixing status, and the realized-volatility regime. A descriptive panel specification is

$$|\Delta_{FB,PM,i,t}| = g\big(\mathrm{Var}_t(\sigma), |m_{i,t}|, \tau_{i,t}, RV_t, L_{i,t}\big) + u_{i,t}, \tag{40}$$

where $L_{i,t}$ contains liquidity diagnostics. The purpose of this regression is mechanism description, not a causal interpretation of market volatility.

4.6 | Primary empirical comparisons

The first empirical question is whether full posterior integration reduces pricing error relative to the posterior-mean, MAP, and MLE plug-ins. The second is whether any difference is concentrated in states predicted by Equation (29). The analysis therefore reports both average performance and conditional performance across maturity, moneyness, volatility regime, option type, and fraction of the averaging month already fixed.

Statistical comparisons will use paired option-date errors because all pricing rules are evaluated on the same market observations and, where possible, the same Monte Carlo random numbers. Standard errors will account for dependence within contracts and dates in the final specification. The exact clustering rule will be fixed before examining the final model-ranking results.

5 | PRELIMINARY RESULTS

This section distinguishes completed evidence from analyses that remain to be run. The simulation results reported below are preliminary validation results for the baseline case $\sigma_0 = 0.25$, $S_0 = K = 100$, and $T = 1$. The empirical data audit is complete for the downloaded Barchart histories. Contract-level WTI APO model prices are not yet reported in this internal draft because the historical first-nearby CL curve and realized-fixing reconstruction must be finalized before a market-pricing comparison is economically meaningful.

5.1 | Synthetic evidence: integration versus plug-in

Table 1 reports the initial repeated-sampling comparison. With only 63 daily observations, full Bayes has RMSE 0.469, compared with 0.469 for the posterior-mean plug-in, 0.487 for MAP, and 0.470 for MLE. The full-Bayes and posterior-mean prices are effectively indistinguishable in both bias and absolute error. As the historical sample increases to 252 and 1,260 observations, all four methods converge and the difference between full Bayes and posterior-mean plug-in becomes numerically negligible.

The near equality of $\widehat{C}_{FB}$ and $\widehat{C}_{PM}$ is economically informative. In this baseline setting, posterior dispersion and curvature are not large enough for Jensen-type effects to generate a meaningful pricing difference. This is consistent with the mechanism in Equation (29): posterior integration is not expected to produce a premium merely because the analysis is Bayesian. The result also accords with the broader observation that parameter-uncertainty effects may be modest in sufficiently informative samples Rombouts and Stentoft (2014).

MAP performs somewhat less stably in the smallest sample, but the ranking is not monotone at every sample size. The evidence therefore does not support a generic claim that one plug-in estimator is uniformly dominated. The production experiment expands volatility, moneyness, and maturity specifically to identify where the pricing map becomes nonlinear enough for posterior integration to separate from plug-in valuation.

5.2 | Posterior price uncertainty in a baseline case

For the corrected baseline implementation, the high-precision risk-neutral benchmark is approximately 6.42. A representative posterior fit gives a full-Bayes value near 6.24, a posterior-mean plug-in near 6.24, a MAP plug-in near 6.22, and an MLE plug-in near 6.23. The corresponding central 95% posterior price interval is approximately $[5.80, 6.73]$.

The interval has a different interpretation from the point-estimator comparison. Even when C_{FB} and C_{PM} are almost identical, the push-forward posterior in Equation (19) can remain non-degenerate. Thus, “integration does not change the point price much” does not imply “parameter uncertainty is zero.” The empirical section will distinguish these two questions by reporting both pricing-rule differences and posterior price dispersion.

5.3 | Empirical data coverage

The Barchart ingestion audit produces the sample summarized in Table 2. The broad maturity coverage is useful, but the cross section becomes thin at the longest horizon. This is a feature of the observed market data rather than a reason to manufacture a balanced panel.

Trading activity is sparse for many contracts. Across several expiry/type groups, only a small percentage of dates report positive volume even when open interest is positive. This makes transaction-price language inappropriate for the full panel. The market benchmark is instead an end-of-day settlement or mark, subject to validation against CME records and sensitivity checks that focus on more actively held contracts.

The maturity groups cover September, October, and November 2026; March and September 2027; March and September 2028; and June 2029. The June 2029 group is particularly sparse, with only a few identified contracts in the current download. Those observations will be retained in the raw panel and may be shown as a long-horizon case while the main statistical conclusions are checked without them.

5.4 | Empirical pricing tests to be completed

The principal empirical results table will compare market-pricing errors for full Bayes, posterior-mean plug-in, MAP, and MLE after the first-nearby futures reconstruction is finalized. The table will report the number of option-date observations, mean error, MAE, RMSE, and paired differences in absolute error. It will be produced for the full main sample and separately for calls and puts.

TABLE 1 Preliminary repeated-sampling pricing performance. The target is the risk-neutral price evaluated at the true volatility, not a posterior-generated target.

N	Method	Bias	MAE	RMSE	95% coverage
63	Full Bayes	-0.0011	0.3543	0.4686	0.9500
63	Posterior-mean plug-in	-0.0012	0.3543	0.4687	0.9500
63	MAP plug-in	-0.0919	0.3783	0.4867	0.9500
63	MLE plug-in	-0.0407	0.3576	0.4701	0.9500
252	Full Bayes	0.0139	0.1988	0.2454	0.9625
252	Posterior-mean plug-in	0.0139	0.1988	0.2454	0.9625
252	MAP plug-in	-0.0212	0.1974	0.2446	0.9625
252	MLE plug-in	0.0035	0.1992	0.2444	0.9625
1260	Full Bayes	-0.0048	0.1010	0.1243	0.9250
1260	Posterior-mean plug-in	-0.0048	0.1010	0.1243	0.9250
1260	MAP plug-in	-0.0128	0.1060	0.1302	0.9250
1260	MLE plug-in	-0.0067	0.1013	0.1243	0.9250

TABLE 2 Current WTI APO data audit.

Item	Count
Raw CSV histories	214
Unique option contracts	213
Option-date observations	11,179
Expiry months represented	8
Folder/expiry mismatches	1

A second set of results will condition on the mechanism variables. The key test is not simply whether

$$\overline{|e^{FB}|} < \overline{|e^{PM}|}, \tag{41}$$

but whether any advantage is concentrated where theory predicts it should be: high posterior variance, high realized-volatility regimes, longer maturity, near-the-money nonlinear exposure, and early stages of the averaging window. If full Bayes remains economically indistinguishable from the posterior-mean plug-in even in these states, that is itself a relevant result for practitioners because the computationally simpler plug-in rule would be difficult to reject.

5.5 | Planned figures

The final manuscript will include four core figures. The first plots the WTI price level and 21-day realized volatility, with regime boundaries defined mechanically rather than by geopolitical labels. The second shows selected contracts that move through out-of-the-money, near-the-money, and in-the-money states over time. The third plots $C^{Q}(\sigma)$ and the posterior density of σ for representative short- and long-maturity contracts, making Equation (29) visually interpretable. The fourth plots $|\Delta_{FB,PM}|$ against posterior variance and maturity, with separate markers or panels for volatility regimes.

No empirical model-ranking claim is made in this draft until those calculations are completed using the contract-consistent first-nearby reconstruction.

6 | ROBUSTNESS AND EXTENSIONS

The baseline model is intentionally simple enough that the source of any full-Bayes/plugin difference can be identified. The final paper will therefore separate robustness to statistical choices from robustness to the commodity-pricing model.

6.1 | Prior and estimation-window sensitivity

The inverse-gamma volatility prior in Equation (13) is a baseline rather than a universal recommendation. The main posterior-pricing comparisons will be repeated with weaker and alternative positive-support priors. A result attributed to parameter uncertainty should not be driven by one prior parameterization. Likewise, empirical posteriors will be estimated under alternative historical windows. Shorter windows react more quickly to volatility changes but contain less information; longer windows concentrate the posterior while potentially mixing distinct volatility regimes. This trade-off is directly relevant to Equation (29).

The MCMC implementation will be checked with multiple chains, effective sample sizes, and convergence diagnostics. Because MAP estimates depend on parameterization, the MAP plug-in is treated as a conventional comparator rather than a uniquely Bayesian decision rule. Posterior integration and posterior-mean plug-in are invariant to the arbitrary choice of reporting a marginal KDE mode.

6.2 | Liquidity and settlement quality

The primary data-quality robustness checks progressively restrict the sample by open interest, remove minimum-tick observations, and distinguish dates with and without positive volume. The purpose is not to assume that zero-volume settlements are unusable, but to determine whether the model ranking changes when attention is restricted to contracts with stronger evidence of active market participation.

Contract histories will also be cross-checked against CME daily settlements on a larger set of dates and strikes where public bulletin observations can be matched. Any systematic discrepancy between Barchart's end-of-day field and CME settlement would require redefining the empirical benchmark before model comparisons are interpreted.

6.3 | Alternative moneyness definitions

The baseline moneyness variable uses the expected final APO average in Equation (8). Robustness checks will compare it with simpler forward-based measures and with normalized intrinsic-distance measures that explicitly condition on the realized portion of the monthly average. The economic conclusions should not depend on a single arbitrary ATM bandwidth.

6.4 | Futures-curve dynamics

Equation (20) is a one-factor benchmark. Crude-oil futures are exposed to term-structure movements that need not be perfectly correlated across maturities. A natural extension replaces the common-volatility representation with a low-dimensional correlated futures-curve model. A further extension allows stochastic volatility or mean reversion. These models can improve absolute fit, but they introduce additional parameters whose uncertainty must also be propagated. They therefore provide a demanding test of whether the qualitative conclusion from the baseline model survives richer dynamics.

The paper will distinguish two claims. The first concerns whether a simple constant-volatility model prices WTI APOs accurately in absolute terms. The second concerns whether integrating parameter uncertainty improves on using point estimates *within the same model.* The second comparison is the main contribution and remains meaningful even when the first reveals model misspecification.

6.5 | Monte Carlo accuracy

All competing pricing rules will be evaluated with common random numbers and sufficiently large path counts so that pairwise differences are not dominated by simulation noise. Selected contracts will be repriced at substantially larger simulation budgets to verify convergence. Where a geometric control variate is not directly available for the exact first-nearby structure, the variance-reduction method will be documented separately from the standard Asian benchmark rather than applying an inappropriate closed form.

6.6 | Sample composition

The principal results will be repeated after excluding the sparsest far-maturity contracts and after balancing maturity buckets as far as the observed data permit. Calls and puts will be reported jointly and separately. Because the downloaded panel was intentionally not chosen only by current open interest, we will also verify that results are not driven by contracts that became economically irrelevant after the sharp change in WTI prices.

7 | CONCLUSION

This paper studies a narrow but practically important question in derivative valuation: when volatility is estimated rather than known, does integrating its posterior distribution materially change the value of a path-dependent commodity option relative to conventional plug-in pricing? The framework enforces a separation that is essential for answering this question correctly. Historical parameters are inferred under the physical measure, while derivative values are computed under the risk-neutral measure. In the baseline Black–Scholes setting, uncertainty about the physical drift is therefore not pricing uncertainty.

The synthetic evidence available at this stage establishes an important benchmark. In regular settings, full posterior integration and pricing at the posterior mean of volatility can be almost indistinguishable. This outcome follows naturally when posterior dispersion is limited and the option-pricing function is approximately linear over the relevant posterior support. MAP pricing is somewhat less stable in the smallest preliminary samples, while MLE remains competitive. These results argue against presenting Bayesian integration as mechanically superior. Its economic value must instead be demonstrated in states where posterior uncertainty interacts with sufficient pricing curvature.

WTI Average Price Options provide a useful empirical test because both elements can vary substantially. Crude-oil volatility changes over time, the futures curve determines the remaining first-nearby fixings, and the exposure of an average-price option changes as its calendar-month average becomes progressively realized. The assembled Barchart panel contains 213 unique calls and puts and 11,179 option-date observations over eight expiry months ranging from September 2026 to June 2029. The breadth of this panel allows the analysis to move beyond one illustrative strike and examine the joint roles of maturity, moneyness, liquidity, partial fixing, and volatility regime.

The next empirical step is intentionally mechanical rather than discretionary: reconstruct the historical first-nearby CL fixing schedule and term structure, estimate rolling posteriors using only information available at each valuation date, and compare all pricing rules against validated end-of-day market settlements. This contract-level reconstruction must precede any claim about empirical model performance. Using a generic continuous futures series or a single fixed CL contract would change the underlying payoff and could confound model error with contract misspecification.

The paper's ultimate contribution will therefore be conditional rather than universal. If posterior integration improves pricing primarily in high-volatility, long-maturity, or highly nonlinear states, the result identifies where the extra computation is economically justified. If the posterior-mean plug-in remains effectively equivalent even in those states, the result places a useful empirical bound on the value of parameter integration for this class of commodity derivatives. Either conclusion is informative for practitioners deciding how much statistical uncertainty must be carried into a pricing engine.

DATA AVAILABILITY

The option histories used in the empirical application were obtained through a commercial Barchart subscription. Redistribution of raw proprietary files is not assumed to be permitted. The public repository therefore prioritizes code, schemas, provenance records, diagnostics, hashes, and derived results consistent with the applicable data license.

CODE AVAILABILITY

The research code is organized as an installable Python package, `bayesian_asian_options`, with documented APIs, tests, deterministic seeds, and reproducible experiment entry points. A versioned code snapshot will accompany the final article.

CONFLICT OF INTEREST

The author declares no conflict of interest.

ACKNOWLEDGMENTS

To be completed before submission.

REFERENCES

Black, F. & Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637–654. doi:10.1086/260062.

Curran, M. (1994) Valuing asian and portfolio options by conditioning on the geometric mean price. *Management Science*, 40(12), 1705–1711.

Guidolin, M. & Timmermann, A. (2003) Option prices under bayesian learning: Implied volatility dynamics and predictive densities. *Journal of Economic Dynamics and Control*, 27(5), 717–769. doi:10.1016/S0165-1889(01)00069-0.

Hastings, W.K. (1970) Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1), 97–109. doi:10.1093/biomet/57.1.97.

Kahalé, N. (2020) General multilevel monte carlo methods for pricing discretely monitored asian options. *European Journal of Operational Research*, 287(2), 739–748. doi:10.1016/j.ejor.2020.04.022.

Kemna, A.G.Z. & Vorst, A.C.F. (1990) A pricing method for options based on average asset values. *Journal of Banking & Finance*, 14(1), 113–129. doi:10.1016/0378-4266(90)90039-5.

Merton, R.C. (1973) Theory of rational option pricing. *The Bell Journal of Economics and Management Science*, 4(1), 141–183. doi:10.2307/3003143.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E. (1953) Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087–1092. doi:10.1063/1.1699114.

Rombouts, J.V.K. & Stentoft, L. (2014) Bayesian option pricing using mixed normal heteroskedasticity models. *Computational Statistics & Data Analysis*, 76, 588–605. doi:10.1016/j.csda.2013.06.023.