*Corresponding Author's Email: tokmakbahri@gmail.com
Proceedings of the International Academic Conference on Management and Economics
Vol. 4, Issue. 1, 2026, pp. 1-27
DOI: https://doi.org/10.33422/conferenceme.v4i1.1857

ISSN: 3030-069X online

Diamond Scientific Publishing

Improving the Willow Tree Method: An Enhanced Variant and an Empirical Comparison with the Binomial Tree Method

Bahri Tokmak* **and Ömür Uğur**
Middle East Technical University, Institute of Applied Mathematics, Çankaya, Turkey

Abstract

This paper revisits the Willow Tree (WT) approach for pricing European call options under geometric Brownian motion and compares its numerical performance with the Black–Scholes benchmark and the Cox–Ross–Rubinstein (CRR) binomial tree. We implement two alternative WT sampling schemes, the first partial moment (FPM) and kurtosis matching (KM) methods, and examine the trade-off between pricing accuracy and computational cost. In low-volatility settings, a moderately refined WT discretization already delivers very small pricing errors at low marginal computational cost, with the FPM sampler providing the most stable performance across strikes. However, under long maturities and higher volatilities, the baseline WT exhibits a systematic negative bias. We show that this deterioration is closely associated with a mismatch in the discrete exponential-martingale condition induced by the bounded support of the terminal WT approximation. To mitigate this effect, we introduce an Esscher reweighting of the terminal WT probabilities that restores the martingale condition for the terminal approximation and substantially reduces pricing errors in the difficult regimes considered. In our experiments, the FPM sampler consistently outperforms KM in terms of residual pricing error, while Esscher reweighting yields the largest gains when maturity and volatility are high. Overall, the results suggest that WT with FPM performs well in benign regimes, whereas the Esscher-corrected version is more reliable when the baseline discretization becomes biased.

Keywords: Option pricing, Willow Tree approximation, Esscher transform (exponential tilting), Binomial lattice approximation

1. Introduction

Recombining lattice methods remain attractive in option pricing because they are transparent, flexible, and readily adapted to early exercise and discrete contractual features. In addition to the classical Cox–Ross–Rubinstein (CRR) binomial tree, Curran's Willow Tree (WT) provides a recombining discretization designed to approximate Brownian motion more efficiently by allocating spatial resolution where probability mass is concentrated. Subsequent work by Xu, Hong, and Qin refined the sampling stage of the WT construction and showed that alternative discrete normal approximations can materially improve pricing accuracy when the number of spatial nodes is modest (Cox et al., 1979; Curran, 2001; Xu et al., 2013).

Since these foundational contributions, the literature on WT has expanded well beyond the original geometric Brownian motion (GBM) setting. Extensions now include American options under jump-diffusion dynamics, moving-average barrier options, one-factor short-rate models, and variable-annuity / guaranteed minimum withdrawal benefit valuation (Xu & Yin, 2014; Lu et al., 2017; Wang & Xu, 2018; Dong et al., 2019; Dong et al., 2020). More recent work has also developed WT methods for American and exotic options under stochastic volatility, Lévy-process models, implied willow trees, joint S&P 500/VIX calibration, and risk-minimization hedging under GARCH models (Ma et al., 2020; Ma et al., 2021; Zhou & Xu, 2023; Dong et al., 2024; Dong et al., 2025; Ma et al., 2025). This broader literature positions the WT not merely as a specialized GBM lattice, but as a flexible numerical framework that has been adapted to a wide range of pricing and risk-management problems.

This paper revisits the Willow Tree (WT) method introduced by Curran (Curran, 2001) in a deliberately simple setting: European call pricing under geometric Brownian motion. Because the Black–Scholes formula is available in closed form, this setting provides a clean benchmark for isolating WT discretization error without the additional complications of early exercise or path dependence (Black & Scholes, 1973). It also allows a transparent comparison with the Cox–Ross–Rubinstein (CRR) lattice, a standard reference among recombining tree methods (Cox et al., 1979). Within this framework, we document a systematic pattern in the baseline WT approximation: it performs well in low-volatility, moderate-maturity regimes, but develops a persistent negative bias as maturity and volatility increase. Our numerical evidence suggests that an important contributor to this deterioration is a mismatch in the discrete exponential-martingale condition for the terminal WT approximation.

To mitigate this effect, we introduce an Esscher reweighting of the terminal WT probabilities. The correction leaves the terminal node locations unchanged and adjusts only their probabilities so that the terminal discrete approximation satisfies the martingale condition at maturity. For European options, this yields a simple, numerically stable modification that does not alter the underlying node construction. We study this correction under the two sampling schemes proposed by Xu, Hong, and Qin (Xu et al., 2013)—first partial moment (FPM) and kurtosis matching (KM)—and compare its accuracy and computational behavior with those of the baseline WT and the CRR binomial tree. The Esscher transform is classical in actuarial and financial mathematics and is used here as a reweighting device for the terminal WT distribution rather than as a new measure-change principle (Esscher, 1932; Gerber & Shiu, 1994).

The contribution of the paper is threefold. First, we document how the pricing error of the baseline WT varies with maturity, volatility, and spatial resolution in European call pricing under GBM. Second, we propose an Esscher correction for the terminal WT distribution and describe a robust numerical procedure for computing the tilting parameter. Third, we provide an empirical comparison across discretization choices, contrasting WT with CRR and

evaluating the relative performance of FPM and KM both before and after Esscher reweighting.

The remainder of the paper is organized as follows. Section 2 reviews the Black–Scholes model, the CRR binomial tree, and the WT construction used in our implementation. Section 3 presents the numerical experiments, including the baseline WT–CRR comparison, resolution sensitivity, and maturity–volatility sensitivity under alternative sampling schemes. Section 4 introduces the Esscher-corrected terminal WT approximation and reports its effect on martingale diagnostics, pricing errors, and robustness near the admissibility boundary. Section 5 discusses practical implications for dividend-paying equities, desk calibration, and dynamic extensions. Section 6 concludes and outlines directions for further work.

2. Review of Mathematical Models

Because the focus of this paper is the Willow Tree approximation, we summarize only the benchmark formulations used in numerical comparisons.

2.1 Black–Scholes–Merton (BSM)

Under the risk-neutral measure, the stock price (S_t) is assumed to follow the geometric Brownian motion:

$$dS_t = (r - q)S_t dt + \sigma S_t dW_t^{\mathbb{Q}},\ \sigma > 0 \tag{1}$$

where r is the continuously compounded risk-free interest rate, q is the continuous dividend yield, and $W_t^{\mathbb{Q}}$ is a standard Brownian motion under the risk-neutral measure. The value $C = C(S,t)$ of a European call option with strike K and maturity T then satisfies the Black–Scholes–Merton partial differential equation:

$$C_t + \frac{1}{2}\sigma^2 S^2 C_{SS} + (r - q)SC_S - rC = 0,\ C(S,T) = (S - K)^+, \tag{2}$$

where subscripts denote partial derivatives. Let $\tau = T - t$. The corresponding closed-form call price is

$$C(S,t) = Se^{-q\tau}\Phi(d_1) - Ke^{-r\tau}\Phi(d_2), \tag{3}$$

where

$$d_1 = \frac{\ln(S/K) + \left(r - q + \frac{1}{2}\sigma^2\right)\tau}{\sigma\sqrt{\tau}},\ d_2 = d_1 - \sigma\sqrt{\tau}, \tag{4}$$

and Φ denotes the standard normal cumulative distribution function. This formula serves as the theoretical benchmark in our numerical experiments (Black & Scholes, 1973; Merton, 1973).

2.2 Cox–Ross–Rubinstein Binomial Tree

The Cox–Ross–Rubinstein (CRR) model approximates the stock price process by a recombining multiplicative tree in discrete time (Cox et al., 1979). Let N denote the number of time steps and $\Delta t = T/N$. Over each step, the stock price moves to either an up state or a down state:

$$S_{t+\Delta t} \in \{uS_t,\ dS_t\},\ 0 < d < u. \tag{5}$$

In the CRR parametrization,

$$u = e^{\sigma\sqrt{\Delta t}},\ d = e^{-\sigma\sqrt{\Delta t}} = \frac{1}{u}, \tag{6}$$

and no-arbitrage requires $0 < d < e^{(r-q)\Delta t} < u$.

The corresponding risk-neutral probability is

$$p = \frac{e^{(r-q)\Delta t} - d}{u-d},\ 1-p = \frac{u-e^{(r-q)\Delta t}}{u-d} \tag{7}$$

After time step n, the stock price at node i is

$$S_{n,i} = S_0\, u^{\,i}\, d^{\,n-i},\ i = 0,1,\ldots,n. \tag{8}$$

For a European payoff H, terminal values are initialized by

$$V_{N,i} = H\big(S_{n,i}\big), \tag{9}$$

and the option value is obtained by backward induction:

$$V_{n,i} = e^{-r\Delta t}\big[pV_{n+1,i+1} + (1-p)V_{n+1,i}\big],\ n = N-1,\ldots,0. \tag{10}$$

For American or Bermudan contracts, continuation values are compared with immediate exercise at the admissible exercise dates:

$$V_{n,i} = \max\Big\{H\big(S_{n,i}\big), e^{-r\Delta t}\big[pV_{n+1,i+1} + (1-p)V_{n+1,i}\big]\Big\}. \tag{11}$$

As $N \to \infty$, the CRR price converges to the Black–Scholes–Merton value for standard European options. In this paper, the CRR tree is used as a benchmark lattice method against which the Willow Tree approximation is compared.

2.3 Willow Tree

The Willow Tree (WT), introduced by Curran (Curran, 2001), is a recombining lattice designed to approximate Brownian motion efficiently by concentrating nodes in high-probability regions and using fewer nodes in the tails. Unlike classical binomial or trinomial trees, whose spatial range expands linearly with time, the WT maintains a fixed-width recombining structure. Although the full WT framework supports rollback for American, Bermudan, and path-dependent claims, the European pricing experiments in this paper use only the terminal discrete distribution. Following Curran and Xu et al. (Xu et al., 2013), the WT is constructed from a discrete approximation to the standard normal, with transition probabilities chosen to preserve martingale, variance, and marginal-weight conditions.

Step 1: Discretization of the standard normal.

Let $(z_i, q_i)_{i=1}^{m}$ be a discrete approximation to $N(0,1)$. Under Curran's original equal-probability quantile construction,

$$z_i = \Phi^{-1}\left(\frac{i-\frac{1}{2}}{m}\right),\ q_i = \frac{1}{m},\ i = 1,\ldots,m, \tag{12}$$

where Φ denotes the standard normal cumulative distribution function as before. The discrete law is required to satisfy

$$\sum_{i=1}^{m} q_i z_i = 0,\ \sum_{i=1}^{m} q_i z_i^2 = 1. \tag{13}$$

With the equal-probability grid in Eq. (12), the mean condition is satisfied by symmetry, while the variance may deviate slightly from one. A simple correction is to adjust the endpoints so that Eq. (13) holds exactly.

Step 2: Fixed spatial nodes and time grid.

Let $0 = t_0 < t_1 < \cdots < t_n = T$ be the time grid and denote $h_k = t_{k+1} - t_k$. At time t_k, the WT node values for the Brownian driver are

$$Y_i(t_k) = \sqrt{t_k} z_i,\ i = 1, \dots, m. \tag{14}$$

A Markov chain is then constructed on these nodes through transition matrices $P^{(k)} = \left(p_{ij}^{(k)}\right)$ with

$$p_{ij}^{(k)} = \mathbb{P}\left(Y(t_{k+1}) = Y_j(t_{k+1}) \,\middle|\, Y(t_k) = Y_i(t_k)\right),\ k = 0, \dots, n-1.$$

The target marginal distribution at each time level is determined by the same weight vector $q = (q_1, \dots, q_m)$.

Step 3: Martingale/variance matching and stationarity.

For each row $i = 1, \dots, m$, the transition probabilities are chosen to satisfy

$$\sum_{j=1}^{m} p_{ij}^{(k)} = 1, \tag{15a}$$

$$\sum_{j=1}^{m} p_{ij}^{(k)} Y_j(t_{k+1}) = Y_i(t_k), \tag{15b}$$

$$\sum_{j=1}^{m} p_{ij}^{(k)} Y_j^2(t_{k+1}) - Y_i^2(t_k) = h_k, \tag{15c}$$

Eq. (15b) enforces the discrete martingale condition for the Brownian driver, while Eq. (15c) matches the intended conditional variance growth. In addition, the unconditional weights are required to remain invariant across time levels:

$$\sum_{i=1}^{m} q_i p_{ij}^{(k)} = q_j,\ j = 1, \dots, m. \tag{16}$$

Step 4: Transition matrix via a constrained program.

For each level k, the matrix $P^{(k)} = \left(p_{ij}^{(k)}\right)$ is obtained by solving a constrained optimization problem of the form

$$\min_{P^{(k)}} \left(\sum_{i=1}^{m} \sum_{j=1}^{m} p_{ij}^{(k)} \left|Y_j(t_{k+1}) - Y_i(t_k)\right|^3\right), \tag{17}$$

subject to Eqs. (15a)–(16) and

$$p_{ij}^{(k)} \geq 0 \text{ for all } i, j. \tag{18}$$

This yields a recombining Markov chain with a fixed number m of spatial nodes at each time level.

Once the Brownian nodes and transition matrices are available, they are mapped into stock-price nodes under risk-neutral GBM:

$$S_i(t_k) = S_0 \exp\left(\left(r - q - \frac{1}{2}\sigma^2\right)t_k + \sigma\sqrt{t_k}z_i\right),\ i = 1, \dots, m,\ k = 0, \dots, n \tag{19}$$

Let $V(i, t_k)$ denote the option value at node (i, k) and define the discount factor

$$d(t_k, t_{k+1}) = e^{-r(t_{k+1} - t_k)}.$$

For contracts requiring the full tree dynamics, valuation proceeds by backward induction:

$$V(i, t_k) = d(t_k, t_{k+1}) \sum_{j=1}^{m} p_{ij}^{(k)} V(j, t_{k+1}),\ k = n-1, \dots, 0,\ i = 1, \dots, m. \tag{20}$$

For American or Bermudan contracts, early exercise is enforced by

$$V(i, t_k) \leftarrow \max\{\text{payoff}(S_i(t_k)),\ V(i, t_k)\}. \tag{21}$$

For the European payoffs considered in this paper, the time-zero value is instead computed directly from the terminal discrete distribution:

$$V_0 = e^{-rT}\mathbb{E}[f(S_T)] \approx e^{-rT} \sum_{i=1}^{m} q_i f(S_i(T)), \tag{22}$$

that is,

$$V_0 = e^{-rT} \sum_{i=1}^{m} q_i f\left(S_0 \exp\left(\left(r - q - \frac{1}{2}\sigma^2\right)T + \sigma\sqrt{T}z_i\right)\right). \tag{23}$$

Here f denotes the payoff function. Thus, for the European experiments reported below, pricing depends only on the terminal WT approximation $(z_i, q_i)_{i=1}^{m}$; the transition matrices are reviewed for completeness and for the broader WT framework, but they are not needed in the actual European valuation step used in this paper.

For fixed lattice width m and time grid t_k, the WT ingredients—nodes z_i, weights q_i, and transition matrices $P^{(k)}$—can be precomputed and stored. Subsequent valuations then reuse these objects, which keeps the per-contract online computation light.

When m is small, the basic endpoint-adjusted quantile sampling may underrepresent the tails of the standard normal distribution, which can degrade accuracy, especially for out-of-the-money payoffs. This motivates the refined sampling strategies of Xu, Hong, and Qin (Xu et al., 2013), who propose two alternatives used in this paper: kurtosis matching and first partial moment matching.

Kurtosis matching sampling: The aim is to construct a discrete normal approximation $(z_i, q_i)_{i=1}^{m}$ that matches mean and variance and also brings the fourth moment close to that of the standard normal. Let $\gamma \in (0,1)$ be a shape parameter, and define provisional symmetric weights by

$$\tilde{q}_i = \frac{\left(i - \frac{1}{2}\right)^{\gamma}}{m},\ i = 1, \dots, \frac{m}{2},$$

together with $\tilde{q}_{m+1-i} = \tilde{q}_i$, then normalize:

$$q_i = \frac{\tilde{q}_i}{\sum_{j=1}^{m} \tilde{q}_j}, \; i = 1, \ldots, m.$$

Let $Q_i = \sum_{j=1}^{i} q_j$, define breakpoints

$$Z_i = \Phi^{-1}(Q_i), \; Z_0 = -\infty, \; Z_m = \infty,$$

and choose nodes $z_i \in [Z_{i-1}, Z_i]$ subject to symmetry $z_{m+1-i} = -z_i$. The nodes are then selected so that

$$\sum_{i=1}^{m} q_i z_i = 0, \; \sum_{i=1}^{m} q_i z_i^2 = 1,$$

while minimizing the kurtosis mismatch

$$\min_{\{z_i\}} \left(\sum_{i=1}^{m} q_i z_i^4 - 3\right)^2. \tag{24}$$

First partial moment sampling: The FPM construction uses the same symmetric probability-design step but chooses the nodes so as to match the first partial moments of the standard normal distribution. Using the same breakpoints Z_l, a convenient aggregate loss can be written as

$$\min_{\{z_i\}} \sum_{l=1}^{m-1} \left| \sum_{j=1}^{m} q_j \left(z_j - Z_l\right)^+ - \int_{-\infty}^{\infty} (z - Z_l)^+ \phi(z) dz \right|, \tag{25}$$

subject to

$$\sum_{i=1}^{m} q_i z_i = 0, \; \sum_{i=1}^{m} q_i z_i^2 = 1, \; Z_{i-1} \le z_i \le Z_i.$$

The integral in Eq. (25) admits the closed form

$$\int_{-\infty}^{\infty} (z - a)^+ \phi(z) dz = \phi(a) - a\left(1 - \Phi(a)\right), \tag{26}$$

Where ϕ is the standard normal density. Matching first partial moments aligns the discrete grid with a family of call-like payoffs and is therefore particularly natural in option-pricing applications.

When non-uniform probabilities q_i are used, the WT transport objective in Eq. (17) is naturally reweighted as

$$\min_{P^{(k)}} \left(\sum_{i=1}^{m} q_i \sum_{j=1}^{m} p_{ij}^{(k)} \left| Y_j(t_{k+1}) - Y_i(t_k) \right|^3 \right) \tag{27}$$

with the constraints in Eqs. (15a)–(18) unchanged.

Because KM and FPM emphasize different aspects of the terminal distribution, we do not impose a priori dominance between them at the modeling stage. Instead, their relative performance is assessed empirically in Section 3 in terms of pricing accuracy, residual bias, and computational cost.

3. Numerical Results and Comparative Analysis

We evaluate European call prices obtained from the Willow Tree (WT) approximation against the Black–Scholes benchmark. For each contract and strike K, our primary diagnostic is the signed relative error

$$\mathrm{RelErr}_{\mathrm{WT}}(K) = \frac{\mathrm{Price}_{\mathrm{WT}}(K) - \mathrm{Price}_{\mathrm{Theo}}(K)}{\mathrm{Price}_{\mathrm{Theo}}(K)} \times 100\%. \tag{28}$$

Signed relative error is used to preserve information on systematic overpricing or underpricing. For the European contracts studied here, WT prices are computed directly from the terminal discrete distribution determined by the sampling scheme and node count m; hence the experiments isolate the quality of the terminal WT approximation under GBM rather than full rollback performance. Reported WT runtimes refer to marginal per-contract evaluation times after any offline precomputation.

As a baseline, we also compute Cox–Ross–Rubinstein (CRR) binomial prices and report the analogous signed relative error

$$\mathrm{RelErr}_{\mathrm{CRR}}(K) = \frac{\mathrm{Price}_{\mathrm{CRR}}(K;N) - \mathrm{Price}_{\mathrm{Theo}}(K)}{\mathrm{Price}_{\mathrm{Theo}}(K)} \times 100\%. \tag{29}$$

Here N denotes the number of time steps. For European options, the CRR prices are computed from the terminal binomial distribution, with N held fixed across strikes so that the same lattice is used throughout each strike sweep. Unless otherwise stated, we report $\mathrm{RelErr}(\cdot)$ against moneyness K/S_0 together with runtime summaries under the same hardware environment. We first compare WT with CRR, then study sensitivity to the node count m, and finally examine the effects of maturity T and volatility σ under kurtosis-matching and first-partial-moment sampling.

3.1 Computational Setup

All numerical experiments were implemented in Python 3.12.4 using SciPy 1.13.1 (Virtanen et al., 2020). SciPy was used in the sampling routines, particularly for normal distribution functions and constrained optimization, and in the Esscher calibration routine for one-dimensional root finding via Brent's method. All computations were performed in double precision using float64 arithmetic. The experiments were run under Windows 11 on an AMD64 machine with 20 CPU cores and 64 GB RAM.

A key practical advantage of the Willow Tree method is the separation between offline construction and online valuation (Curran, 2001; Xu et al., 2013). In the full WT framework, the transition matrices in Eq. (17), subject to Eqs. (15a)–(18), and the KM/FPM sampling grids in Eqs. (24)–(26) can be precomputed and reused for a given choice of m, γ, sampling scheme, and time grid. In the European experiments considered here, the online valuation step uses only the terminal WT quadrature in Eqs. (22)–(23); hence, reported marginal WT runtimes refer to pricing on the precomputed terminal grid rather than to the offline construction stage.

3.2 Willow vs. Binomial: Baseline Comparison

We first benchmark the CRR lattice against the Black–Scholes reference and then compare it with the WT approximation. Unless stated otherwise, WT prices are computed on a fixed

terminal grid with $m = 180$ nodes using either the first-partial-moment or kurtosis-matching sampler. Common model parameters are $S_0 = 100$, $r = 0.05$, $q = 0$, $T = 10$ years, and $K \in [80,180]$, with volatility levels $\sigma \in \{0.30,0.50,0.70\}$. For the CRR benchmark, we use $N \in \{50,100,200,300,400,500\}$ steps with the standard parametrization $u = e^{\sigma\sqrt{\Delta t}}$, $d = 1/u$, and $p = (e^{(r-q)\Delta t} - d)/(u - d)$, and report the signed relative error in Eq. (29). Keeping N fixed across strikes ensures a like-for-like comparison of error profiles and runtimes.

Figure 1 reports CRR results in the low-volatility regime ($\sigma = 0.3$). As N increases, the signed relative error shrinks toward zero at both strikes: for $K = 100$, it decreases from -0.2942% at $N = 50$ to -0.0295% at $N = 500$; for $K = 180$, it decreases from -0.2466% to -0.0604%. The associated runtime rises from approximately 0.12–0.13 ms to 1.28–1.31 ms over the same range.

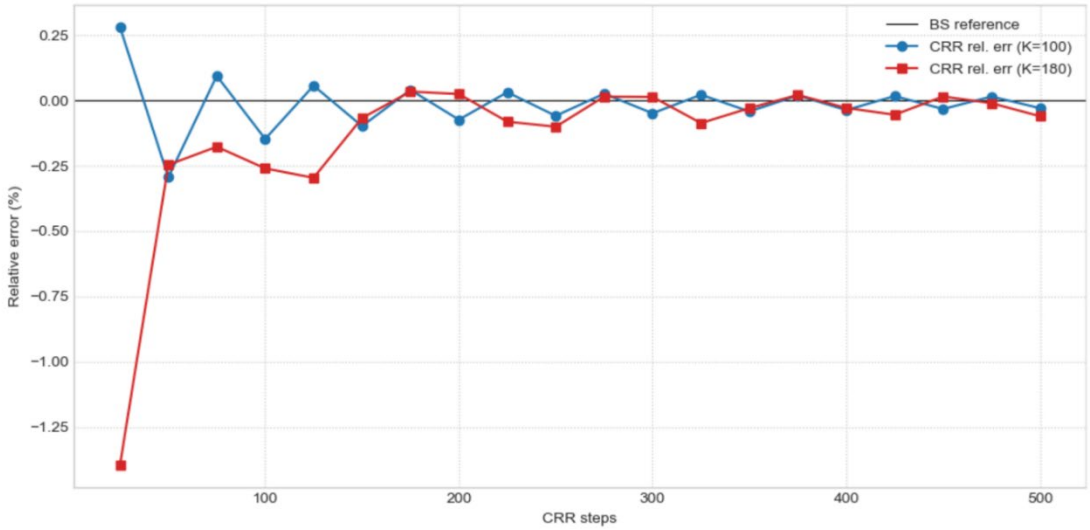

Figure 1. Binomial CRR relative error versus number of steps for K ∈{100, 180}

Source: Prepared by the authors

Figure 2 shows the corresponding WT results under the FPM sampler. In the same low-volatility regime, WT–FPM achieves very small errors at low marginal cost: at $K = 100$, $\mathrm{RelErr}_{\mathrm{WT}} = -0.0016\%$ with a runtime of about 0.007 ms, and at $K = 180$, the error is -0.0020% with essentially the same runtime. Across strikes, the error remains mildly negative and exhibits a saw-tooth pattern, with local peaks at the node-aligned strikes marked by red stars. In this regime, FPM provides a stable and highly accurate terminal approximation.

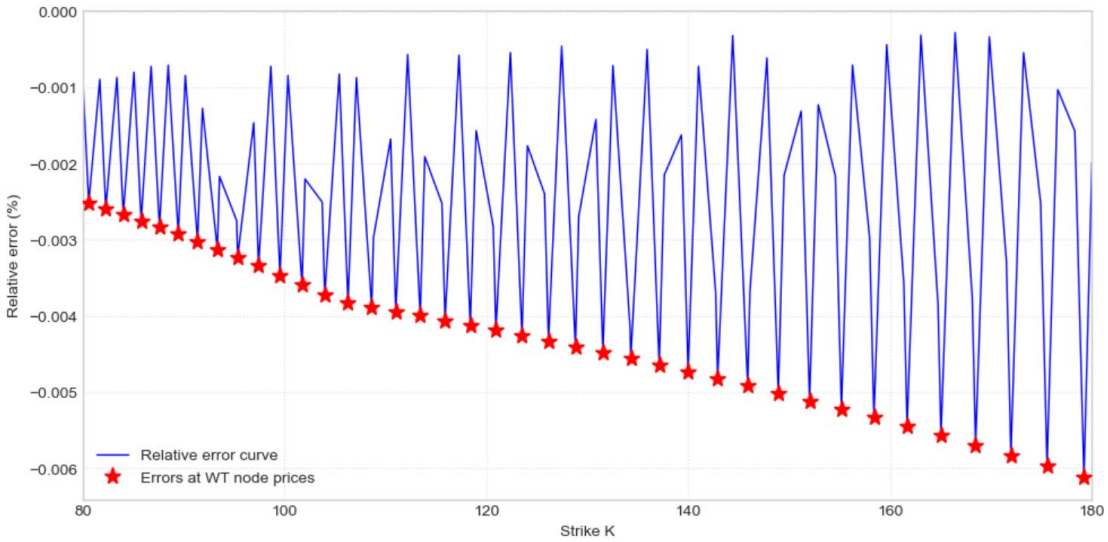

Figure 2. WT (FPM) relative error versus strike $K \in [80, 180]$

Source: Prepared by the authors

Figure 3 reports the corresponding WT results for the KM sampler. In the low-volatility regime, the errors remain negative, indicating underpricing, while runtime stays at about 0.007–0.008ms. At $K = 100$, the error is -0.3419%, markedly worse than CRR at $N = 500$; at $K = 180$, it is -0.0200%, improving on CRR's -0.0604%. Thus, KM performs reasonably well in the deeper OTM region but is less accurate near the money than FPM.

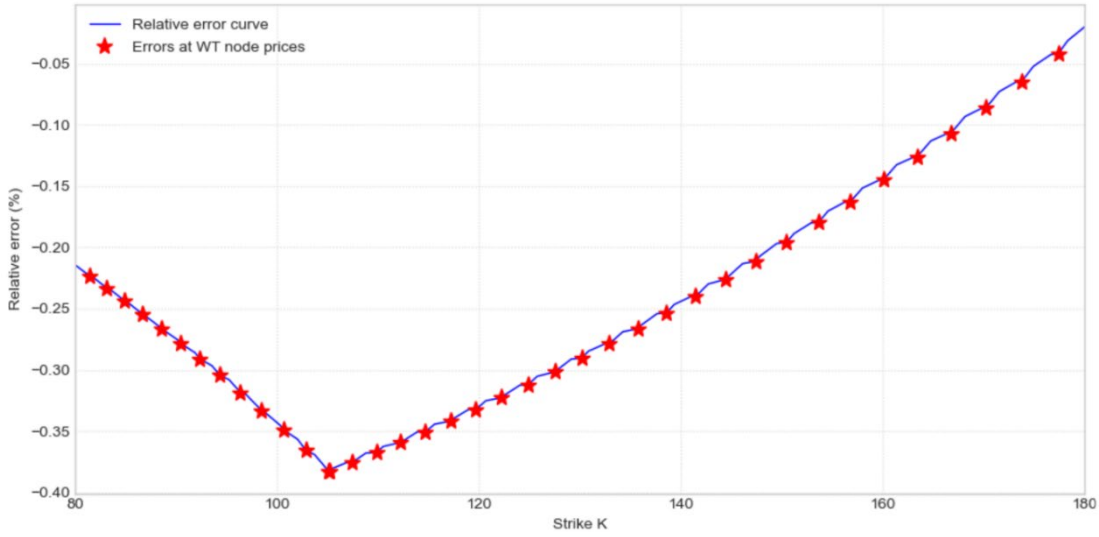

Figure 3. WT (KM) relative error versus strike $K \in [80, 180]$

Source: Prepared by the authors

3.3 Resolution Sensitivity of the Willow Tree Approximation

We next examine how WT accuracy depends on the spatial resolution of the terminal discretization. Since the European prices considered here are computed directly from the terminal WT distribution, this experiment should be interpreted as a sensitivity study with respect to the node count m. We fix $(S_0, r, q, T, \sigma) = (100, 0.05, 0, 10, 0.3)$ and study the

signed relative error as a function of m for two strikes, $K \in \{100,180\}$, under each sampling scheme.

Under the first-partial-moment sampler, the error decreases rapidly as the grid is refined. As shown in Figure 4 for $K = 100$, the signed relative error falls from about -4.9% at $m = 10$ to approximately $+0.004\%$ at $m = 300$, with the curve already essentially flat by $m \approx 180$. The higher-strike case $K = 180$ exhibits the same pattern, with the error dropping from about -6.5% on very coarse grids to approximately $+0.006\%$ at the finest reported resolution. This indicates fast and stable resolution convergence of the FPM terminal approximation across both moneyness regimes.

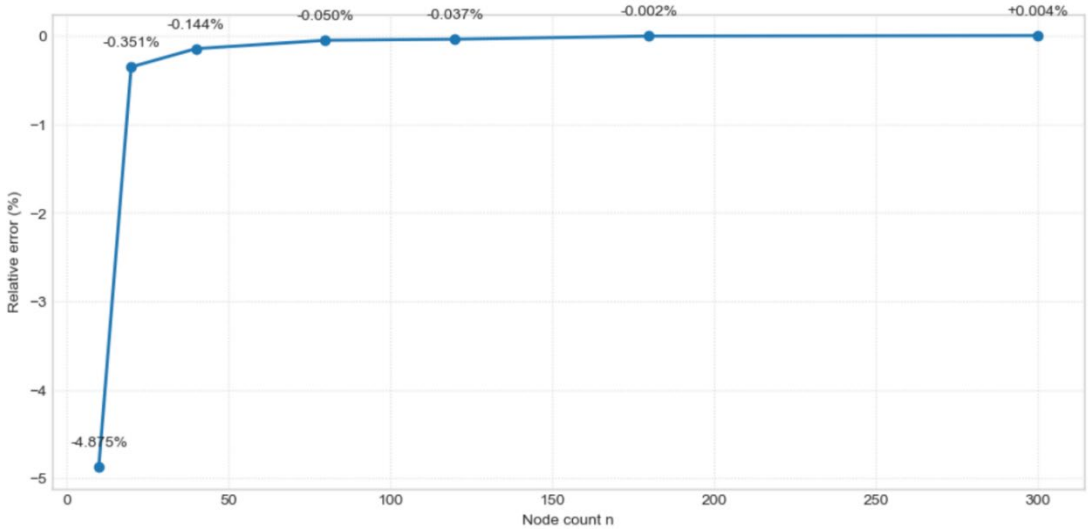

Figure 4. WT (FPM) relative error vs. node count (σ = 0.3, K = 100, T = 10)

Source: Prepared by the authors

The kurtosis-matching sampler shows the same broad improvement with increasing m, but with larger and less stable residuals. Figure 5 shows that for $K = 180$ the error settles near -0.039% at $m = 300$, after a mildly non-monotone path. The near-the-money case remains less accurate: the error initially overshoots positive on coarse grids and then converges to about -0.219% at $m = 300$. Overall, WT accuracy is highly sensitive to very coarse discretization but stabilizes rapidly with refinement. In the present experiments, $m \approx 180$ provides a practically effective operating point: under FPM the signed relative error is already in the $\mathcal{O}(10^{-3})\%$ range, whereas under KM the error also stabilizes but at a materially higher residual level.

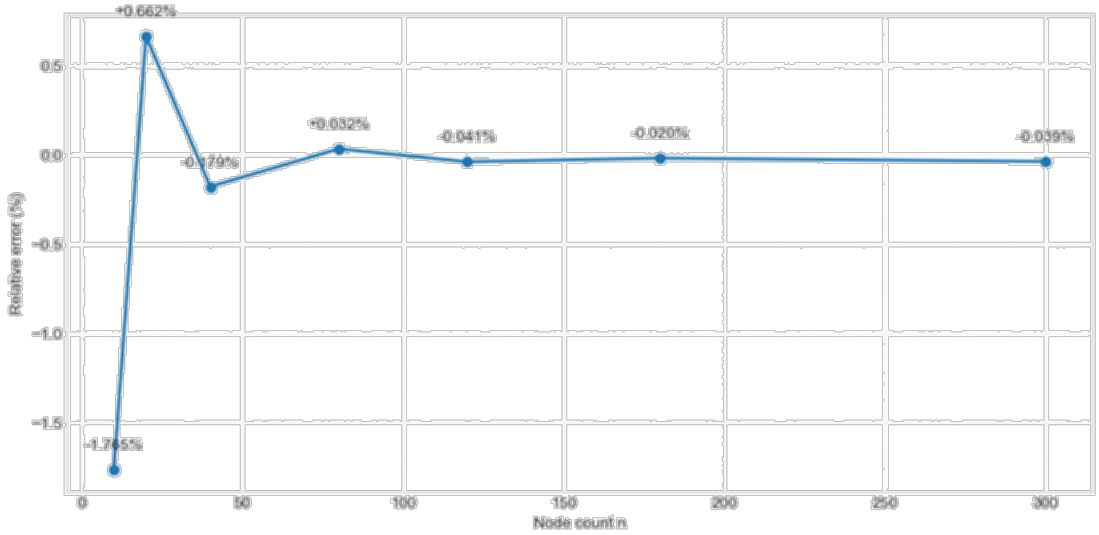

Figure 5. WT (KM) relative error vs. node count (σ = 0.3, K = 180, T = 10)

Source: Prepared by the authors

3.4 Sampling Sensitivities: Maturity and Volatility Regimes

We now examine how WT performance changes across maturity and volatility. Since the mechanism discussed later depends on the scale parameter $\sigma\sqrt{T}$, we vary T and σ separately while holding the remaining model inputs fixed. Throughout this subsection, $S_0 = 100$, $r = 0.05$, and $q = 0$; results are reported for $K \in \{100,180\}$, $T \in \{1,2,\dots,20\}$ and $\sigma \in \{0.5,0.7\}$. Unless otherwise stated, the WT terminal discretization uses $m = 180$ nodes and $\gamma = 2/3$.

Figure 6 shows the results for the first-partial-moment sampler. The pricing bias remains negative and grows with both maturity and volatility. For $K = 100$, the error reaches -1.8423% at $T = 20$ when $\sigma = 0.5$, and -14.6725% when $\sigma = 0.7$. The $K = 180$ case is similar, with terminal errors of -2.0152% and -15.1857%, respectively. These results indicate a clear weakness of the baseline WT terminal approximation in long-maturity, high-volatility regimes.

A similar but more severe pattern is observed under kurtosis matching. At $T = 20$, the $K = 100$ errors are -7.9448% for $\sigma = 0.5$ and -32.6533% for $\sigma = 0.7$, while the corresponding $K = 180$ errors are -8.4709% and -33.6269%. Thus, although KM can behave reasonably in shorter-maturity, low-volatility settings, its negative bias grows substantially faster than under FPM as maturity and volatility increase. In the regimes considered here, FPM is therefore the more robust of the two baseline samplers.

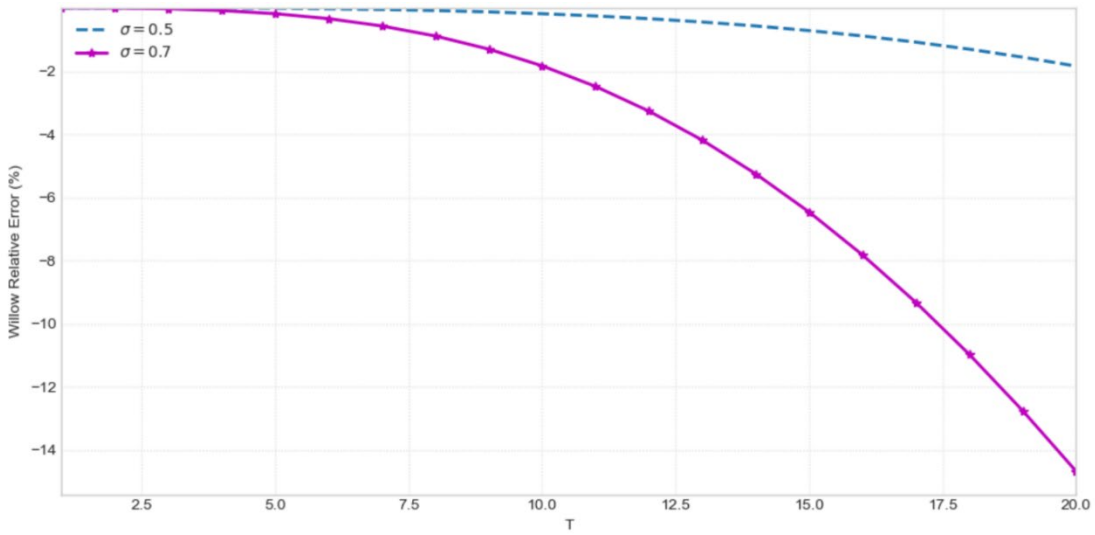

Figure 6. WT with first partial moment sampling relative error as a function of maturity T

Source: Prepared by the authors

4. Analysis of Willow Tree Deficiencies and a Practical Enhancement

To explain the deterioration documented in Section 3, it is useful to compare the exact exponential-martingale identity under risk-neutral GBM with the identity implied by the terminal WT approximation used in our European pricing experiments. For simplicity, we state the argument for a non-dividend-paying stock ($q = 0$).

Under risk-neutral GBM,

$$S_T = S_0 \exp\left(\left(r - \frac{1}{2}\sigma^2\right)T + \sigma W_T\right), W_T \sim N(0,T),$$

and therefore, the discounted stock price is a martingale so that

$$e^{-rT}\mathbb{E}[S_T] = S_0.$$

Equivalently,

$$\mathrm{E}\left[e^{\sigma W_T - \frac{1}{2}\sigma^2 T}\right] = 1.$$

This identity is simply the terminal manifestation of the martingale property of the discounted stock price under GBM.

In the European WT approximation used in this paper, $W_T/\sqrt{T}$ is replaced by a discrete proxy Z supported on $\{z_i\}_{i=1}^m$ with probabilities $\{q_i\}_{i=1}^m$. Writing

$$\alpha := \sigma\sqrt{T},$$

define

$$b(\sigma,T) := e^{-\alpha^2/2}\sum_{i=1}^m q_i e^{\alpha z_i} = e^{-\frac{1}{2}\sigma^2 T}\sum_{i=1}^m q_i e^{\sigma\sqrt{T} z_i}.$$

Under the terminal WT approximation,

$$e^{-rT}\,\mathbb{E}_{\mathrm{WT}}[S_T] = S_0\,b(\sigma,T).$$

Hence the exact martingale identity is preserved at maturity if and only if $b(\sigma,T) = 1$.

The following observation identifies a regime in which the bounded support of the WT terminal approximation necessarily pushes the exponential moment below its GBM benchmark.

Proposition 1. Let Z take values $z_1,\dots,z_m$ with probabilities $q_1,\dots,q_m$ where $q_i \geq 0$, $\sum_{i=1}^{m} q_i = 1$, and the support is bounded above by

$$z_{max} := max_{1\leq i\leq m} z_i < \infty.$$

Then

$$b(\sigma,T) \leq exp\left(-{}^{1}\!/_{2}\,\sigma^2 T + \sigma\sqrt{T} z_{max}\right), \tag{30}$$

and, consequently,

$$b(\sigma,T) < 1 \text{ whenever } \sigma\sqrt{T} > 2z_{max}. \tag{31}$$

Moreover,

$$\forall T > 0:\ b(\sigma,T) \to 0\ as\ \sigma \to \infty,\ \forall \sigma > 0:\ b(\sigma,T) \to 0\ as\ T \to \infty. \tag{32}$$

Proof. Since $q_i \geq 0$ and $\sum_i q_i = 1$,

$$\sum_{i=1}^{m} q_i e^{\sigma\sqrt{T} z_i} \leq \left(\max_i e^{\sigma\sqrt{T} z_i}\right)\sum_{i=1}^{m} q_i = e^{\sigma\sqrt{T} z_{max}},$$

which yields Eq. (30). Writing $\alpha = \sigma\sqrt{T}$, Eq. (30) becomes

$$b(\sigma,T) \leq exp\left(-{}^{1}\!/_{2}\,\alpha^2 + \alpha z_{max}\right).$$

If $\alpha > 2z_{max}$, then

$$-{}^{1}\!/_{2}\,\alpha^2 + \alpha z_{max} = \alpha\left(-{}^{1}\!/_{2}\,\alpha + z_{max}\right) < 0,$$

and hence, the right-hand side is strictly smaller than one, proving Eq. (31). Finally, completing the square in Eq. (30) gives

$$-{}^{1}\!/_{2}\,\sigma^2 T + \sigma\sqrt{T} z_{max} = -{}^{1}\!/_{2}\left(\sigma\sqrt{T} - z_{max}\right)^2 + {}^{1}\!/_{2}\,z_{max}^2,$$

hence

$$b(\sigma,T) \leq exp\left({}^{1}\!/_{2}\,z_{max}^2\right) exp\left(-{}^{1}\!/_{2}\left(\sigma\sqrt{T} - z_{max}\right)^2\right).$$

The right-hand side tends to zero when $T > 0$ is fixed and $\sigma \to \infty$, or when $\sigma > 0$ is fixed and $T \to \infty$. This proves Eq. (32).

The proposition shows that bounded support does not force $b(\sigma,T) < 1$ for every (σ,T). Particularly, for small values of $\sigma\sqrt{T}$, the sign of $b(\sigma,T) - 1$ depends on the higher-moment structure of the discrete approximation, as discussed in Appendix A. What the proposition establishes is a regime result: once $\sigma\sqrt{T}$ becomes sufficiently large relative to $z_{\max}$, the terminal WT approximation understates the exponential martingale. Since

$$e^{-rT}\,\mathbb{E}_{\mathrm{WT}}[S_T] = S_0\,b(\sigma,T),$$

this provides a natural mechanism for the persistent negative bias observed in long-maturity and high-volatility settings, although it does not imply that every European call price is underestimated in all parameter configurations.

The same mechanism is visible numerically. Figure 7 plot

$$Y(t) \coloneqq e^{-\frac{1}{2}\sigma^2 t}\sum_{i=1}^{m} q_i e^{\sigma\sqrt{t}z_i}$$

for $m = 180$ and volatilities $\sigma \in \{0.3, 0.5, 0.7\}$, with the unit line representing the exact GBM identity. Under the FPM sampler, $Y(t)$ remains close to unity for $\sigma = 0.3$, drifts only mildly below one for $\sigma = 0.5$, and declines more visibly for $\sigma = 0.7$ as t increases. This behavior is consistent with the upper bound in Eq. (30) and with the regime condition in Eq. (31). The KM sampler shows a more pronounced decline in the challenging regimes, consistent with the larger pricing bias in Section 3 and the higher-order moment deficits reported in Appendix A. These observations motivate the Esscher correction introduced next, which reweights the terminal WT probabilities to restore the martingale condition at maturity.

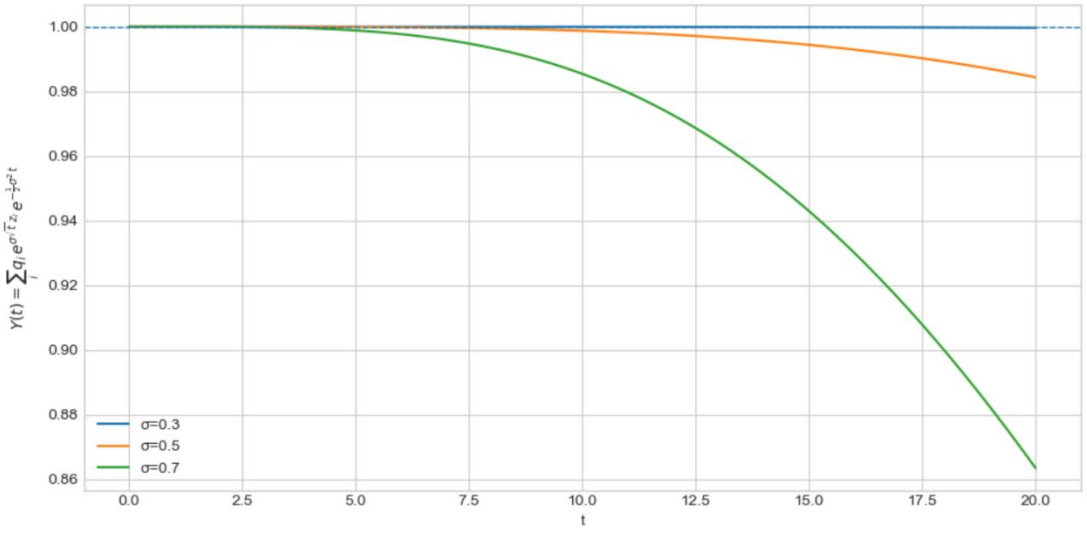

Figure 7. WT approximation of the exponential martingale under the FPM sampler

Source: Prepared by the authors

4.1 Esscher (Exponential-Tilting) Correction for the Terminal WT Approximation

A practical way to mitigate the exponential-martingale defect identified above is to reweight the terminal WT probabilities by an Esscher transform, while keeping the terminal node locations fixed. This is the appropriate correction in the present paper, because European prices are computed from the terminal WT distribution rather than from a full rollback over WT transition matrices.

Let Z denote the discrete WT proxy supported on $\{z_i\}_{i=1}^m$ with base probabilities $q_i{}_{i=1}^m$, where $q_i > 0$ and $\sum_{i=1}^m q_i = 1$. Define the discrete moment generating function (mgf) and cumulant generating function (cgf) by

$$M(\theta) = \sum_{i=1}^m q_i e^{\theta z_i},\ K(\theta) = logM(\theta).$$

For any $\theta \in \mathbb{R}$, the Esscher–tilted probabilities are

$$q_i^{(\theta)} = \frac{q_i e^{\theta z_i}}{\sum_{j=1}^m q_j e^{\theta z_j}} = \frac{q_i e^{\theta z_i}}{M(\theta)},\ i = 1, \dots, m. \tag{33}$$

These tilted probabilities define a new law on the same support. Under this law, the tilted mgf is

$$M_\theta(u) \coloneqq E_\theta[e^{uZ}] = \sum_{i=1}^m q_i^{(\theta)} e^{uz_i} = \sum_{i=1}^m \frac{q_i e^{\theta z_i}}{M(\theta)} e^{uz_i} = \frac{1}{M(\theta)} \sum_{i=1}^m q_i e^{(\theta+u)z_i}$$

$$= \frac{M(\theta+u)}{M(\theta)} = exp(log\, M(\theta+u) - log\, M(\theta)) = exp\big(K(\theta+u) - K(\theta)\big). \tag{34}$$

The Esscher transform is classical in actuarial and financial mathematics (Esscher, 1932; Gerber & Shiu, 1994), and it can also be interpreted as a minimum-relative-entropy adjustment among reweighting that matches a prescribed exponential moment (Kallsen & Shiryaev, 2002).

Let $\alpha := \sigma\sqrt{T} > 0$. For the terminal WT approximation used in the European pricing step, the risk-neutral martingale condition at maturity is

$$e^{-\alpha^2/2}\mathbb{E}[e^{\alpha Z}] = 1.$$

Using Eq. (34), this condition is equivalent to

$$\mathbb{E}_\theta[e^{\alpha Z}] = e^{\alpha^2/2} \Leftrightarrow K(\theta+\alpha) - K(\theta) = \alpha^2/2. \tag{35}$$

Thus, the Esscher calibration problem reduces to solving Eq. (35) for θ. The numerical implementation of this calibration is summarized in Appendix B.

To study uniqueness and existence, define

$$f(\theta) := K(\theta+\alpha) - K(\theta) - \alpha^2/2.$$

Since

$$K''(\vartheta) = Var_\vartheta(Z) \geq 0,$$

with strict inequality for non-degenerate support, K' is increasing and therefore

$$f'(\theta) = K'(\theta+\alpha) - K'(\theta) = \mathbb{E}_{\theta+\alpha}[Z] - \mathbb{E}_\theta[Z] > 0,\ (\alpha > 0).$$

Hence f is strictly increasing; so Eq. (35) has at most one solution. Existence requires that the target exponential moment remains attainable by reweighting the probabilities on the same fixed support.

Let

$$z_{min} := \min_i z_i, \; z_{max} := \max_i z_i.$$

As $\theta \to -\infty$, the tilted law concentrates on z_{min}, whereas as $\theta \to \infty$, it concentrates on z_{max}. Consequently,

$$\lim_{\theta \to -\infty} [K(\theta + \alpha) - K(\theta)] = \alpha z_{min}, \lim_{\theta \to \infty} [K(\theta + \alpha) - K(\theta)] = \alpha z_{max}.$$

Therefore, a finite solution to Eq. (35) exists if

$$\alpha z_{min} < \frac{\alpha^2}{2} < \alpha z_{max},$$

or equivalently,

$$z_{min} < \frac{\alpha}{2} < z_{max}. \tag{36}$$

For the symmetric WT grids used here, this reduces to the practical admissibility condition

$$\alpha < 2z_{max}. \tag{37}$$

This condition does not contradict the result of the previous subsection. Proposition 1 identified a sufficient regime in which the baseline WT approximation must understate the exponential martingale, namely when $\alpha = \sigma\sqrt{T}$ is sufficiently large relative to z_{max}. By contrast, Eq. (37) characterizes when that defect can still be corrected exactly by reweighting probabilities on the same fixed support. If $\alpha \geq 2z_{max}$, then the support itself is too narrow, and an exact maturity calibration of the form in Eq. (33) is no longer feasible without enlarging or redesigning the terminal grid. In other words, the Esscher correction is effective precisely when the target exponential moment remains reachable on the fixed WT support.

Once the unique solution θ_T of Eq. (35) is obtained, we retain the terminal nodes z_i and replace the original probabilities q_i by the tilted probabilities $q_i^{(\theta_T)}$. Then, by construction,

$$e^{-\alpha^2/2} \sum_{i=1}^{m} q_i^{(\theta_T)} e^{\alpha z_i} = e^{-\alpha^2/2} \frac{M(\theta_T + \alpha)}{M(\theta_T)} = 1. \tag{38}$$

Consequently, for a non–dividend stock,

$$e^{-rT} \mathbb{E}_{\theta_T}[S_T] = e^{-rT} S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)T} \mathbb{E}_{\theta_T}[e^{\alpha Z}] = S_0.$$

Thus, the discounted stock expectation is restored at maturity under the reweighted terminal WT approximation. European payoffs are then evaluated by replacing q_i with $q_i^{(\theta_T)}$ in the terminal WT quadrature while keeping the same nodes. It is important to stress that this is a maturity-specific correction of the terminal WT distribution, not a level-by-level reweighting of WT transition matrices.

To assess numerical stability, one may solve the same calibration problem horizon by horizon. Specifically, for each $t \in [0, T]$, let $\theta(t)$ solve

$$K\big(\theta(t) + \sigma\sqrt{t}\big) - K\big(\theta(t)\big) = \frac{1}{2}\sigma^2 t,$$

and define

$$Y(t) := e^{-\frac{1}{2}\sigma^2 t} \sum_{i=1}^{m} q_i^{(\theta(t))} e^{\sigma\sqrt{t}z_i}.$$

Under exact numerical calibration, $Y(t) = 1$ for each t. Accordingly, Table 1 and Table 2 should be interpreted primarily as diagnostics of the numerical calibration step rather than as separate pricing results. For both samplers and all reported volatility levels, the computed values satisfy $Y(T) = 1.000000$ up to floating-point precision, and the uniform deviations $\max_{t\in[0,T]} |Y(t) - 1|$ remain at the 10^{-13} level. This confirms that the root-finding procedure is numerically stable in the parameter range considered.

Table 1. Esscher tilt diagnostic at maturity (m = 180, algorithm = kurtosis matching)

σ	$Y(T)$	$max_{t\in[0,T]}\|Y(t)-1\|$
0.3	1.000000	1.847411×10^{-13}
0.5	1.000000	2.120526×10^{-13}
0.7	1.000000	2.124967×10^{-13}

Source: Prepared by the authors

Table 2. Esscher tilt diagnostic at maturity (m = 180, algorithm = first partial moment)

σ	$Y(T)$	$max_{t\in[0,T]}\|Y(t)-1\|$
0.3	1.000000	6.550316×10^{-14}
0.5	1.000000	3.326228×10^{-13}
0.7	1.000000	3.279599×10^{-13}

Source: Prepared by the authors

To complement these diagnostics with price level evidence, Tabs. 3 and 4 report the effect of Esscher reweighting on long-maturity European calls with $T = 20$, $m = 180$, $K \in \{100,180\}$, and $\sigma \in \{0.3,0.5,0.7\}$. Each table compares the baseline WT approximation with the maturity-calibrated Esscher correction from Eq. (35). Across both samplers, the correction substantially reduces the negative long-maturity bias. Under FPM (Table 3), errors at $\sigma = 0.7$ fall from -14.673% to -0.531% for $K = 100$, and from -15.095% to -0.825% for $K = 180$. Under KM (Table 4), the corresponding high-volatility errors decrease from -32.653% to -1.453% and from -33.627% to -2.348%. Thus, Esscher reweighting provides a substantial correction of the terminal WT bias whenever the target exponential moment remains attainable on the fixed support, with FPM retaining smaller post-correction residuals than KM.

Table 3. WT relative errors (%) at T = 20; m = 180, algorithm = first partial moment. Columns show None (no tilt) vs Esscher tilt

σ	$K = 100$ (None)	$K = 100$ (Esscher)	$K = 180$ (None)	$K = 180$ (Esscher)
0.3	-0.045%	-0.004%	-0.055%	-0.009%
0.5	-1.842%	-0.113%	-1.983%	-0.196%
0.7	-14.673%	-0.531%	-15.095%	-0.825%

Source: Prepared by the authors

Table 4. WT relative errors (%) at T = 20; m = 180, algorithm = kurtosis matching. Columns show None (no tilt) vs Esscher tilt

σ	$K = 100$ (None)	$K = 100$ (Esscher)	$K = 180$ (None)	$K = 180$ (Esscher)
0.3	-0.732%	-0.174%	-0.733%	-0.120%
0.5	-7.945%	-0.429%	-8.593%	-0.739%
0.7	-32.653%	-1.453%	-33.627%	-2.348%

Source: Prepared by the authors

4.2 Robustness Near the Esscher Admissibility Boundary

The fixed-support feasibility of the Esscher correction can be summarized by the ratio

$$\rho = \frac{\alpha}{2z_{max}} = \frac{\sigma\sqrt{T}}{2z_{max}}.$$

The admissibility condition in Eq. (37) is equivalent to $\rho < 1$. Values of ρ close to one indicate that the target exponential moment is near the largest value attainable on the fixed terminal support. To assess robustness, we consider the FPM sampler under the common inputs $S_0 = 100$, $K = 100$, $r = 0.05$, and $q = 0$, and vary m, γ, σ, and T. The diagnostic quantities reported in Table 5 are z_{max}, the feasibility ratio ρ, feasibility of the fixed-support Esscher calibration, the maximum martingale residual $\max_t |\, Y(t) - 1 \,|$, and the post-correction WT relative error.

Table 5. Robustness diagnostics near the Esscher admissibility boundary for the FPM sampler

Case	m	γ	σ	T	z_{max}	ρ	*Feasible*	*Martingale Residual*	*WT RelErr(%)*
1	180	2/3	0.7	20	3.6702	0.4265	Yes	3.2796×10^{-13}	-0.531
2	120	2/3	0.85	50	3.4461	0.8721	Yes	9.3714×10^{-13}	-2.149
3	120	2/3	0.9	50	3.4461	0.9234	Yes	9.6489×10^{-13}	-2.695
4	120	2/3	0.95	50	3.4461	0.9747	Yes	9.5757×10^{-13}	-3.744
5	120	2/3	1.0	50	3.4461	1.0260	No	4.6868×10^{-1}	-55.072
6	180	2/3	1.0	50	3.6702	0.9633	Yes	7.4685×10^{-13}	-3.158
7	120	1	1.0	50	3.7488	0.9431	Yes	1.2352×10^{-12}	-2.898

Source: Prepared by the authors

These diagnostics show that the main long-maturity/high-volatility case remains safely admissible: in Case 1 of Table 6, $\rho = 0.4265$, with a martingale residual at the 10^{-13} level. Stress tests in Cases 2–4 move closer to the admissibility boundary by using $m = 120$ and larger (σ, T). As ρ increases from 0.8721 to 0.9747, calibration remains feasible and the

martingale residual stays near machine precision, although the post-correction pricing error increases in magnitude. This indicates that the terminal support should not be chosen too close to the feasibility limit.

Case 5 illustrates the failure mode: when $\rho > 1$, the target exponential moment is not attainable on the fixed support, so the correction cannot restore the martingale condition. Cases 6 and 7 show that feasibility can be restored by enlarging the effective right-tail support: either by increasing m from 120 to 180, or by increasing γ to 1. Both adjustments reduce ρ below one and substantially improve the pricing error.

A practical implementation rule is therefore to avoid operating directly at the boundary. For a safety margin $\eta > 0$, one may require

$$z_{max} \geq (1+\eta)\frac{\alpha}{2}, \ \eta \in [0.05, 0.10],$$

before applying the fixed-support Esscher correction. If this condition is violated, the terminal grid should be reconstructed with a larger support. In the present experiments, increasing m or increasing γ both provide effective ways of restoring feasibility.

5. Practical Implications for Dividend-Paying Equities, Desk Calibration, and Dynamic Extensions

For dividend-paying equities represented by a continuous dividend yield q, the terminal Esscher correction requires no structural modification. Under the risk-neutral dynamics

$$dS_t = (r-q)S_t dt + \sigma S_t dW_t,$$

the terminal WT nodes can be written in forward form as

$$S_i(T) = F_{0,T}\exp\left(-\frac{1}{2}\sigma^2 T + \sigma\sqrt{T}z_i\right), \ F_{0,T} = S_0 e^{(r-q)T}.$$

Thus, the dividend yield affects only the deterministic forward level $F_{0,T}$, while the Esscher correction acts on the standardized shock Z. The calibration equation is therefore unchanged:

$$K(\theta_T + \alpha) - K(\theta_T) = \frac{\alpha^2}{2}, \ \alpha = \sigma\sqrt{T},$$

where $K(\theta) = \log M(\theta)$.

From a desk-calibration perspective, the tilting parameter is not directly strike-dependent when a single volatility input is used for a fixed maturity. For maturity or total-variance buckets $l = 1, \ldots, L$, let

$$v_l = \sigma_l^2 T_l.$$

The corresponding Esscher parameter solves

$$K(\theta_l + \sqrt{v_l}) - K(\theta_l) = \frac{v_l}{2}.$$

Once θ_l and the tilted probabilities are computed, they can be stored and reused for all European calls and puts sharing the same terminal WT grid and total-variance input. If strike-dependent implied volatilities are used, then for the κ-th strike $K_{l,\kappa}$ at maturity T_l, the relevant total variance is

$$v_{l,\kappa} = \sigma_{\text{imp}}^2\left(K_{l,\kappa}, T_l\right)T_l,$$

and the same one-dimensional calibration is repeated for each distinct total variance or bucket.

Discrete cash dividends require a different treatment when modeled as explicit payments at known intermediate dates. Then the one-step terminal lognormal representation in Eqs. (22)–(23) is generally insufficient, because the stock process has deterministic jumps before maturity. For European claims, this does not require early-exercise rollback, but it requires multi-step WT propagation. Within the WT framework, dividend dates should be inserted into the time grid; the transition matrices are used between dividend dates, while the dividend event itself is handled by a deterministic stock adjustment and, if necessary, interpolation or remapping. This is consistent with the reusable-lattice logic of WT implementations, where transition objects can be precomputed for relevant time-step structures and reused when the corresponding grid is needed (Curran, 2001; Xu et al., 2013).

For American, Bermudan, or genuinely path-dependent claims, the terminal Esscher correction is not sufficient by itself. These products depend on intermediate values, exercise decisions, or path variables, and therefore require the full WT dynamic structure. Valuation proceeds through the transition matrices $p_{ij}^{(k)}$ as in Eq. (20), together with the early-exercise update in Eq. (21) when applicable. A dynamically consistent Esscher-type extension would need to operate at the transition or time-level level while preserving recombination, non-negativity, row-stochasticity, the WT martingale and variance constraints, and exponential-martingale consistency across time levels. These extensions are beyond the scope of the present European terminal-approximation study.

6. Conclusion and Further Work

This paper revisited the Willow Tree (WT) method for European call pricing under geometric Brownian motion and compared it with the Black–Scholes formula and the Cox–Ross–Rubinstein (CRR) binomial tree. In low-volatility and moderate-maturity regimes, the terminal WT approximation delivered high accuracy at low marginal computational cost, with the first partial moment (FPM) sampler consistently outperforming kurtosis matching (KM) and $m \approx 180$ providing an effective operating point. However, the baseline WT approximation developed a systematic negative bias in long-maturity and high-volatility settings, which was closely associated with a mismatch in the terminal exponential-martingale condition. A maturity-calibrated Esscher reweighting of the terminal WT probabilities substantially reduced this bias under both samplers, with smaller post-correction residuals under FPM than under KM. Overall, the results indicate that baseline WT performs well in benign regimes, while the Esscher-corrected terminal approximation provides an effective improvement when the baseline discretization becomes biased.

Several extensions follow naturally. First, it would be useful to study whether the same conclusions hold across other European payoffs, such as digital, asset-or-nothing, or spread-type contracts, where sensitivity to tail and kink structure may differ from that of plain-

vanilla calls. Second, the present Esscher correction is a maturity-specific adjustment of the terminal WT distribution; extending the same idea to American, Bermudan, or genuinely path-dependent contracts would require a formulation compatible with rollback and dynamic consistency across time levels. A systematic comparison of such extensions against CRR, Monte Carlo, or regression-based early-exercise methods would help clarify the trade-offs among accuracy, bias, and computational cost in more general pricing environments.

Acknowledgment

This research was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under Project No. 124F138. The authors gratefully acknowledge TÜBİTAK for its financial support.

Conflict of Interest

The authors have declared that there is no conflict of interest.

Author Contributions

Both authors contributed equally to this work.

References

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

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of Financial Economics, 7*(3), 229-263. https://doi.org/10.1016/0304-405X(79)90015-1

Curran, M. (2001). Willow Power: Optimizing derivative pricing trees. *Algo Research Quarterly, 4*(4), 15–24.

Dong, B., Xu, W., & Cui, Z. (2024). Implied willow tree. *The Journal of Derivatives, 31*(4), 44–74. https://doi.org/10.3905/jod.2024.1.200

Dong, B., Xu, W., & Cui, Z. (2025). Joint implied willow tree: An approach for joint S&P 500/VIX calibration. *Journal of Futures Markets, 45*(6), 547–568. https://doi.org/10.1002/fut.22572

Dong, B., Xu, W., & Kwok, Y. K. (2019). Willow tree algorithms for pricing guaranteed minimum withdrawal benefits under jump-diffusion and CEV models. *Quantitative Finance, 19*(10), 1741–1761. https://doi.org/10.1080/14697688.2019.1583360

Dong, B., Xu, W., Šević, A., & Šević, Ž. (2020). Efficient willow tree method for variable annuities valuation and risk management. *International Review of Financial Analysis, 68*, 101429. https://doi.org/10.1016/j.irfa.2019.101429

Esscher, F. (1932). On the probability function in the collective theory of risk. *Skandinavisk Aktuarietidskrift, 15*(3), 175-195. https://doi.org/10.1080/03461238.1932.10405883

Gerber, H. U., & Shiu, E. S. W. (1994). Option pricing by Esscher transforms. *Transactions of the Society of Actuaries, 46*, 99-140.

Kallsen, J., & Shiryaev, A. N. (2002). The cumulant process and Esscher's change of measure. *Finance and Stochastics, 6*, 397-428. https://doi.org/10.1007/s007800200069

Lu, L., Xu, W., & Qian, Z. (2017). Efficient willow tree method for European-style and American-style moving average barrier options pricing. *Quantitative Finance, 17*(6), 889–906. https://doi.org/10.1080/14697688.2016.1231416

Ma, J., Huang, S., & Xu, W. (2020). An efficient convergent willow tree method for American and exotic option pricing under stochastic volatility models. *The Journal of Derivatives, 27*(3), 75–98. https://doi.org/10.3905/jod.2019.1.092

Ma, J., Wang, C., & Xu, W. (2025). A new lattice approach for risk-minimization hedging under generalized autoregressive conditional heteroskedasticity models. *European Journal of Operational Research, 321*(3), 1021–1035. https://doi.org/10.1016/j.ejor.2024.10.002

Ma, J., Xu, W., & Yao, Y. (2021). Cosine willow tree structure under Lévy processes with application to pricing variance derivatives. *The Journal of Derivatives, 29*(2), 30–60. https://doi.org/10.3905/jod.2021.1.140

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

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., & SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods, 17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Wang, G., & Xu, W. (2018). A unified willow tree framework for one-factor short-rate models. *The Journal of Derivatives, 25*(3), 33–54. https://doi.org/10.3905/jod.2018.1.064

Xu, W., & Yin, Y. (2014). Pricing American options by willow tree method under jump-diffusion process. *The Journal of Derivatives, 22*(1), 46–56. https://doi.org/10.3905/jod.2014.22.1.046

Xu, W., Hong, Z., & Qin, C. (2013). A new sampling strategy willow tree method with application to path-dependent option pricing. *Quantitative Finance, 13*(6), 861–872. https://doi.org/10.1080/14697688.2012.762111

Zhou, Z., & Xu, W. (2023). Robust willow tree method under Lévy processes. *Journal of Computational and Applied Mathematics, 424*, 114982. https://doi.org/10.1016/j.cam.2022.114982

Appendix

Appendix A: Moment Expansion for the Terminal WT Exponential Term

Let Z denote the terminal WT discrete proxy supported on $z_1, \dots, z_m$ with probabilities $q_1, \dots, q_m$, where $q_i \geq 0, \sum_{i=1}^{m} q_i = 1$, and the discrete law is constructed so that

$$\mathbb{E}[Z] = 0, \ \mathbb{E}[Z^2] = 1.$$

In the applications considered in this paper, the WT samplers are symmetric, so the odd moments vanish:

$$\mathbb{E}[Z^{2k+1}] = 0, \ k \geq 0.$$

For a generic scale parameter $\alpha > 0$ (in the main text, $\alpha = \sigma\sqrt{t}$), define

$$b(\alpha) := e^{-\alpha^2/2}\mathbb{E}[e^{\alpha Z}] = e^{-\alpha^2/2}\sum_{i=1}^{m} q_i e^{\alpha z_i}. \tag{A.1}$$

This is the discrete analogue of the exponential-martingale term under GBM. Let $X \sim N(0,1)$. Since

$$e^{-\alpha^2/2}\mathbb{E}[e^{\alpha X}] = 1,$$

we may write

$$b(\alpha) - 1 = e^{-\alpha^2/2}(\mathbb{E}[e^{\alpha Z}] - \mathbb{E}[e^{\alpha X}]). \tag{A.2}$$

Using the power-series expansion of the exponential function,

$$\mathbb{E}[e^{\alpha Z}] = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}\mathbb{E}[Z^k], \ \mathbb{E}[e^{\alpha X}] = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!}\mathbb{E}[X^k].$$

Substituting into Eq. (A.2) gives

$$b(\alpha) - 1 = e^{-\alpha^2/2}\sum_{k=0}^{\infty} \frac{\alpha^k}{k!}(\mathbb{E}[Z^k] - \mathbb{E}[X^k]). \tag{A.3}$$

Because Z and X are both centered, symmetric, and have unit variance, the differences in the zeroth, first, second, and all odd moments vanish. Hence the expansion reduces to

$$b(\alpha) - 1 = e^{-\alpha^2/2}\sum_{k=2}^{\infty} \frac{\alpha^{2k}}{(2k)!}\Delta_{2k}, \ \Delta_{2k} := \mathbb{E}[Z^{2k}] - (2k-1). \tag{A.4}$$

Eq. (A.4) makes explicit the statement used in Section 4: for small α, the sign of $b(\alpha) - 1$ is governed by the leading nonzero even-moment mismatch between the discrete WT approximation and the Gaussian benchmark. In particular, if $\Delta_4 \neq 0$, then

$$b(\alpha) - 1 = \frac{\Delta_4}{24}\alpha^4 + O(\alpha^6), \ \alpha \to 0. \tag{A.5}$$

On the other hand, if $\Delta_4 = 0$ but $\Delta_6 \neq 0$, then

$$b(\alpha) - 1 = \frac{\Delta_6}{720}\alpha^6 + O(\alpha^8), \ \alpha \to 0. \tag{A.6}$$

Thus, no global conclusion of the form $b(\alpha) < 1$ for all $\alpha \neq 0$ follows from bounded support alone. Rather, small-α behavior depends on which even when moment mismatch appears first and on its sign. This is why Section 4 formulates Proposition 1 as a regime result for sufficiently large α, while the present appendix explains the more local behavior through moment matching.

To quantify these moment mismatches, we report relative errors of the WT even moments ($M_{2k}^{(m)}$) against the Gaussian benchmark (μ_{2k}).

$$M_{2k}^{(m)} = \sum_{i=1}^{m} q_i z_i^{2k}, \mu_{2k} = (2k-1).$$

Define

$$\varepsilon_{2k}^{(m)} = \frac{M_{2k}^{(m)} - \mu_{2k}}{\mu_{2k}} \times 100\%. \quad (A.7)$$

For $m \in \{10,14, \ldots,180\}$, results for both samplers are reported in Tabs. 6 and 7.

Table 6. Relative error (%) table for WT moments under the FPM sampler (rows: m, columns: moment order n)

m	$n=2$	4	6	8	10	12	14	16
10	0.0000	-36.9364	-70.9689	-89.1785	-96.5939	-99.0752	-99.7804	-99.9539
14	0.0000	-2.3732	-17.9529	-43.3959	-67.5757	-84.2225	-93.3468	-97.5301
16	0.0000	-3.5356	-19.6383	-44.0114	-67.0560	-83.3614	-92.6796	-97.1582
20	0.0000	-0.7479	-8.5502	-25.4759	-47.2269	-67.5877	-82.6341	-91.8077
40	0.0000	-0.0458	-3.1313	-13.0017	-29.7683	-49.6248	-67.8972	-81.7283
80	0.0000	-0.0186	-1.5630	-7.8294	-20.7345	-38.6449	-57.5148	-73.6245
120	0.0000	-0.0146	-0.9234	-4.9869	-14.3154	-28.8432	-46.1239	-62.8787
180	0.0000	0.0409	-0.0901	-1.5138	-6.1676	-15.5306	-29.3450	-45.5715

Source: Prepared by the authors

Table 7. Relative error (%) table for WT moments under the KM sampler (rows: m, columns: moment order n)

m	$n=2$	4	6	8	10	12	14	16
10	0.0000	0.0000	-9.2743	-32.9667	-59.7400	-79.8973	-91.4559	-96.8457
14	0.0000	0.0000	-12.5101	-32.8004	-55.8487	-75.3402	-88.1748	-95.0586
16	0.0000	0.0000	-21.8312	-52.5433	-76.7674	-90.5124	-96.6840	-98.9886
20	0.0000	0.0000	-2.5196	-13.0869	-32.3408	-54.6456	-73.7002	-86.6570
40	0.0000	0.0000	-8.0274	-24.6957	-45.5480	-65.1567	-80.1519	-89.8620
80	0.0000	0.0000	-5.7961	-18.7831	-36.6221	-55.2991	-71.4368	-83.4029
120	0.0000	0.0000	-5.3346	-16.9748	-33.2184	-50.9375	-67.0950	-79.8208
180	0.0000	0.0000	-4.3045	-13.9788	-28.0947	-44.4078	-60.3395	-73.8958

Source: Prepared by the authors

The tables show that both samplers match the second moment by construction, while KM also matches the fourth moment almost exactly (Table 6). Beyond these matched moments, higher even moments increasingly fall below the Gaussian benchmark, with slower convergence at higher orders. At $m = 180$, KM's leading mismatch is typically the negative sixth-moment term, whereas FPM has a small positive fourth-moment error ($+0.0409\%$, Table 5). Hence, for small α, the sign of $b(\alpha) - 1$ may differ across samplers. This complements Section 4: small-α behavior is governed by leading even-moment mismatches, while large-α behavior is driven by bounded support, explaining the stronger long-maturity/high-volatility bias and the role of the Esscher correction when the target exponential moment is attainable.

Appendix B: Pseudocode for the Esscher Calibration Routine

The numerical implementation of the Esscher calibration routine is summarized in Table 8.

Table 8. Esscher calibration for the terminal WT distribution

Step	Operation
Input	Terminal WT nodes $\{z_i\}_{i=1}^{m}$; base probabilities $\{q_i\}_{i=1}^{m}$; volatility σ; maturity T; root-finding tolerance ϵ_{root}; residual tolerance ϵ_R; maximum iteration number N_{max}.
Output	Esscher tilting parameter θ_T; tilted probabilities $\{q_i^{(\theta_T)}\}_{i=1}^{m}$; martingale residual R.
1	Set $\alpha = \sigma\sqrt{T}$.
2	If $\alpha = 0$, set $\theta_T = 0$ and $\{q_i^{(\theta_T)}\} = q_i$ for all $i = 1, \dots, m$; return θ_T, $\{q_i^{(\theta_T)}\}$, and $R = 0$.
3	Compute the support endpoints $z_{min} = \min_i z_i$ and $z_{max} = \max_i z_i$.
4	Check the fixed-support feasibility condition $z_{min} < \alpha/2 < z_{max}$. If this condition fails, exact Esscher calibration on the current terminal support is not feasible; enlarge or redesign the terminal grid. In applications, one may impose the stronger safety-margin condition $z_{max} \geq (1+\eta)\frac{\alpha}{2}$, $\eta \in [0.05, 0.10]$ to avoid operating too close to the boundary.
5	Define the cumulant generating function $K(\theta) = \log\left(\sum_{i=1}^{m} q_i e^{\theta z_i}\right)$. For numerical stability, evaluate $K(\theta)$ by a log-sum-exp transformation.
6	Define the scalar calibration function $f(\theta) = K(\theta + \alpha) - K(\theta) - \alpha^2/2$.
7	Find a bracketing interval $[L, U]$ such that $f(L) < 0$ and $f(U) > 0$. Since f is strictly increasing, the admissible calibration has a unique root once a valid bracket is found.
8	Solve $f(\theta_T) = 0$ on $[L, U]$ using bisection or Brent's method until $\lvert f(\theta_T)\rvert < \epsilon_{root}$ or N_{max} is reached.
9	Compute the tilted probabilities $q_i^{(\theta_T)} = \frac{q_i e^{\theta_T z_i}}{\sum_{j=1}^{m} q_j e^{\theta_T z_j}}$, for $i = 1, \dots, m$.

Step	Operation
10	Compute the martingale residual $R = e^{-\frac{\alpha^2}{2}} \sum_{i=1}^{m} q_i^{(\theta_T)} e^{\alpha z_i} - 1$.
11	Accept the calibration if $\mid R \mid < \epsilon_R$. If $\mid R \mid \geq \epsilon_R$, repeat the root solve with a smaller root-finding tolerance $\epsilon_{\mathbf{root}}$, a wider bracket, or a larger $N_{\mathbf{max}}$. If the residual remains large despite convergence, reconstruct the terminal grid with wider support.

Source: Prepared by the authors