Parameter Recovery for the Vasicek Model Using Kolmogorov-Arnold Networks: A Synthetic Study with Zero-Coupon Pricing Implications

Tshewang Phuntsho[1,*], Deyu Meng[2], and Jamyang Tenzin[1]
[1] School of Business Intelligence and Information Systems, Gedu College of Business Studies, Royal University of Bhutan, Bhutan
[2] Department of Information and Communication Sciences, Sophia University, Tokyo, Japan
Email: tshewangphuntsho.gcbs@rub.edu.bt[1,*]; mengdy539@eagle.sophia.ac.jp[2]; jamyangtenzin.gcbs@rub.edu.bt[1]
* Corresponding author

Received: 24 May 2026, Revised: 6 July 2026, Accepted: 8 August 2026, Published: 13 August 2026

Abstract
When governments, corporations, municipalities, or other public institutions require medium- or long-term funding, they frequently issue bonds to investors. This paper studies parameter calibration for the Vasicek short-rate model and the pricing implications of recovered parameters for zero-coupon bonds. The theory is stated under the risk-neutral measure, while the physical-to-risk-neutral mapping is summarized explicitly to avoid ambiguity in the pricing drift. A spline-based Kolmogorov-Arnold Network (KAN) is compared with exact maximum likelihood estimation (EMLE), Euler-Maruyama maximum likelihood estimation (EMMLE), a quadratic-variation-initialized two-stage likelihood baseline (QV-2S), and a multilayer perceptron (MLP). The training corpus contains 1.2 million independent synthetic Vasicek paths, and the common out-of-sample test set contains 10,000 additional paths. KAN obtains the lowest overall root mean squared error (RMSE) for the long-run mean, volatility, and mean-reversion speed, although its advantage over the MLP is small. Downstream results are metric-dependent: KAN achieves the lowest aggregate zero-coupon price RMSE, whereas the MLP achieves a slightly lower aggregate yield RMSE. Paired path-level bootstrap intervals confirm KAN's price-RMSE advantage over all comparators, but also confirm the MLP's small yield-RMSE advantage over KAN. The findings therefore support supervised amortized parameter recovery in this in-distribution synthetic design, but do not establish that KAN uniformly dominates either likelihood estimation or a conventional neural network.
Keywords: *Vasicek model, affine term structures, Kolmogorov-Arnold Networks, synthetic calibration, parameter recovery, zero-coupon bonds*

1. INTRODUCTION

When governments, corporations, municipalities, or other public institutions require medium- or long-term funding, they frequently issue bonds to investors. The prices of such instruments depend on promised cash flows and on the evolution of prevailing interest rates. In the special case of a zero-coupon bond, the holder receives a single payment at maturity, which makes the instrument a natural starting point for term-structure modeling (Vasicek, 1977; Brigo & Mercurio, 2006). Bond prices and yields are linked by an exponential identity, so the short end of the curve is fundamental for fixed-income valuation. The short rate is defined from the local maturity behavior of the term structure and, under no-arbitrage arguments, the specification of short-rate dynamics under the risk-neutral measure determines zero-coupon prices (Vasicek, 1977; Duffie & Kan, 1996; Brigo & Mercurio, 2006).

The Vasicek model is the canonical Gaussian short-rate model (Vasicek, 1977). Its analytical tractability makes it useful for illustrating affine pricing logic and for studying how parameter-estimation error propagates into bond-pricing implications (Brigo & Mercurio, 2006). Implementation nevertheless requires recovery of the structural parameters governing mean reversion, the long-run level, and volatility. Classical approaches include exact likelihood estimation, discretization-based likelihood methods, and realized-volatility-style procedures for diffusion models (Kloeden & Platen, 1992; Yoshida, 1992; Phillips & Yu, 2009; Barndorff-Nielsen & Shephard, 2002; Aït-Sahalia, 2002). In parallel, Kolmogorov-Arnold Networks (KANs) have recently been proposed as spline-based learning architectures that approximate nonlinear maps with learnable univariate functions on network edges (Liu et al., 2024; Bodner et al., 2024).

The objective of this paper is threefold. First, it states the Vasicek zero-coupon pricing framework under risk-neutral parameters and illustrates price and yield sensitivity. Second, it evaluates whether a spline-based KAN can recover Vasicek parameters from engineered path features under a controlled synthetic design. Third, it examines how parameter-recovery errors propagate into model-implied zero-coupon prices and yields. EMLE, EMMLE, and QV-2S provide model-based benchmarks, while a matched-width MLP distinguishes a possible KAN-specific gain from the broader effect of supervised amortized estimation.

We do not claim that KAN replaces exact likelihood estimation when the Vasicek transition density is known. Instead, we evaluate whether a feature-based KAN can learn an amortized inverse map from simulated short-rate paths to risk-neutral Vasicek parameters under a controlled synthetic prior. The evidence is interpreted jointly across parameter, price, and yield metrics because our experiment does not identify one method as uniformly best.

The remainder of the paper is organized as follows. Section 2 presents the Vasicek bond-pricing framework and pricing sensitivities. Section 3 summarizes the benchmark recovery procedures. Section 4 presents the experimental design and results. Section 5 concludes and identifies the limitations of the synthetic evidence.

2. VASICEK MODEL AND PRICING IMPLICATIONS

For a zero-coupon bond maturing at time T, let $P(t,T)$ denote its time-t price and let $R(t,T)$ denote the continuously compounded spot rate. The zero-coupon bond price is

$$P(t,T) = \exp\left[-(T-t)R(t,T)\right], \tag{1}$$

and the short rate is

$$r_t = \lim_{\Delta t \to 0} R(t, t+\Delta t) = -\left.\frac{\partial}{\partial T} \ln P(t,T)\right|_{T=t}. \tag{2}$$

Under the risk-neutral measure $\mathbb{Q}$, the Vasicek short rate follows the Gaussian mean-reverting diffusion (Vasicek, 1977; Brigo & Mercurio, 2006)

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t^{\mathbb{Q}}, \qquad \kappa > 0, \quad \sigma > 0. \tag{3}$$

For a zero-coupon bond, the pricing partial differential equation is

$$\begin{cases} \partial_t P(t,T) + \kappa(\theta - r_t)\partial_{r_t} P(t,T) \\ \quad + \dfrac{1}{2}\sigma^2 \partial_{r_t r_t} P(t,T) - r_t P(t,T) = 0, \\ P(T,T) = 1. \end{cases} \tag{4}$$

Affine short-rate models admit the exponential-affine representation (Duffie & Kan, 1996; Brigo & Mercurio, 2006)

$$P(t,T) = \exp\left[A(\tau) - B(\tau)r_t\right], \qquad \tau = T - t. \tag{5}$$

Substitution into the pricing equation gives

$$B'(\tau) = 1 - \kappa B(\tau), \qquad B(0) = 0, \tag{6}$$

$$A'(\tau) = -\kappa\theta B(\tau) + \frac{1}{2}\sigma^2 B(\tau)^2, \qquad A(0) = 0. \tag{7}$$

The closed-form solution is

$$B(\tau) = \frac{1 - e^{-\kappa\tau}}{\kappa}, \tag{8}$$

$$A(\tau) = \left(\theta - \frac{\sigma^2}{2\kappa^2}\right)\left[B(\tau) - \tau\right] - \frac{\sigma^2}{4\kappa} B(\tau)^2. \tag{9}$$

The continuously compounded yield is therefore

$$y(t,T) = -\frac{1}{\tau}\ln P(t,T) = -\frac{A(\tau)}{\tau} + \frac{B(\tau)}{\tau} r_t. \tag{10}$$

As $\tau \to \infty$,

$$B(\tau) \to \frac{1}{\kappa}, \qquad \frac{B(\tau)}{\tau} \to 0, \tag{11}$$

and hence

$$y_\infty = \lim_{\tau \to \infty} y(t,T) = \theta - \frac{\sigma^2}{2\kappa^2}. \tag{12}$$

The economically relevant long-maturity restriction concerns the sign of y_∞. When $y_\infty > 0$, the zero-coupon bond price decays toward zero with maturity. When $y_\infty \le 0$, the Gaussian Vasicek model implies a non-positive long-end yield, which may be economically implausible unless a negative-rate regime is intentionally allowed (Vasicek, 1977; Brigo & Mercurio, 2006).

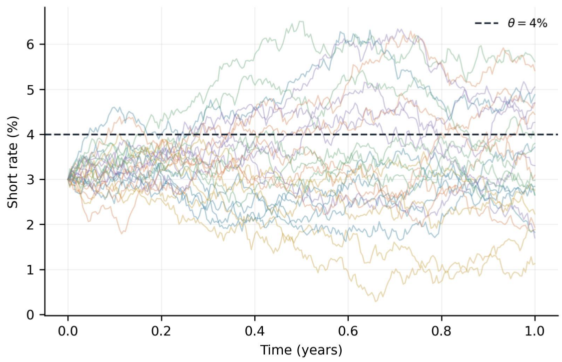

Fig. 1: Simulated Vasicek short-rate paths for $\theta = 4\%$, $\sigma = 1.5\%$, $\kappa = 0.5$, and $r_0 = 3\%$.

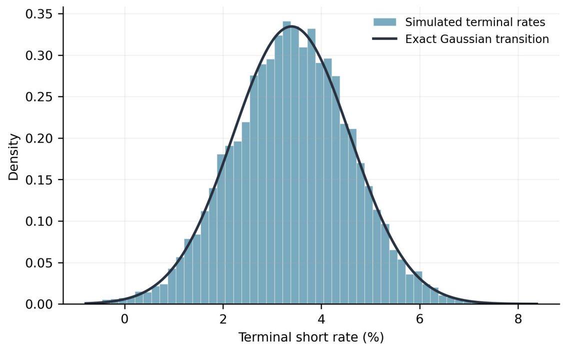

Fig. 2: Empirical distribution of simulated terminal short rates and the corresponding exact Gaussian transition density.

For fixed maturity $\tau > 0$, bond prices are strictly decreasing in the current short rate because

$$\frac{\partial P(t,T)}{\partial r_t} = -B(\tau)P(t,T) < 0. \tag{13}$$

If the model is first written under the physical measure $\mathbb{P}$, a market-price-of-risk specification is required to recover the pricing drift. Under a constant market price of risk λ, one convenient formulation is

$$dr_t = \kappa_{\mathbb{P}}(\theta_{\mathbb{P}} - r_t)dt + \sigma dW_t^{\mathbb{P}}, \tag{14}$$

$$dW_t^{\mathbb{Q}} = dW_t^{\mathbb{P}} + \lambda dt, \qquad \theta_{\mathbb{Q}} = \theta_{\mathbb{P}} - \frac{\sigma\lambda}{\kappa_{\mathbb{P}}}. \tag{15}$$

In the pricing formulas above, θ denotes the risk-neutral long-run mean. This separation matters because historical-path estimation and risk-neutral pricing are not interchangeable without an explicit bridge between $\mathbb{P}$ and $\mathbb{Q}$ (Vasicek, 1977; Duffie & Kan, 1996; Brigo & Mercurio, 2006).

Figures 1 and 2 illustrate mean reversion and the exact Gaussian transition. Figures 3–6 show zero-coupon price sensitivities. Figures 7 and 8 show a representative yield curve and its evolution with the observed state.

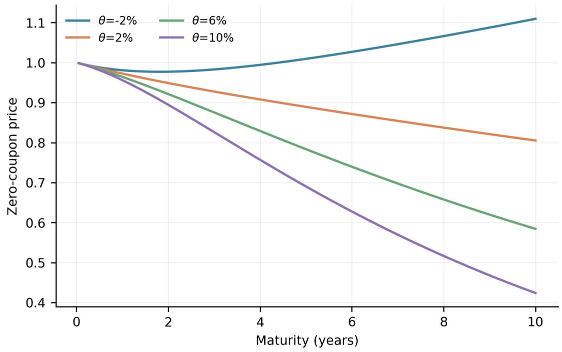

Fig. 3: Sensitivity of zero-coupon bond prices to the risk-neutral long-run mean θ.

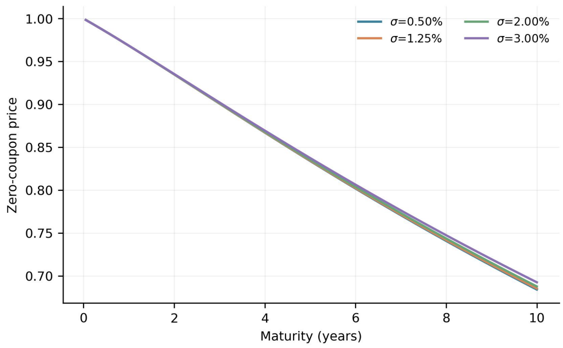

Fig. 4: Sensitivity of zero-coupon bond prices to volatility σ.

3. BENCHMARK RISK-NEUTRAL PARAMETER-RECOVERY PROCEDURES

The short rate is simulated under $\mathbb{Q}$ and observed at equally spaced times $t_0, t_1, \ldots, t_n$ with step size Δt. Let

$$\phi^{\mathbb{Q}} = (\theta^{\mathbb{Q}}, \sigma, \kappa^{\mathbb{Q}}) \tag{16}$$

denote the risk-neutral parameter vector. The objective is to recover $\phi^{\mathbb{Q}}$ from a simulated short-rate path and assess how the estimates propagate into zero-coupon pricing.

3.1 Exact Maximum Likelihood Estimation

EMLE provides a natural benchmark because the Vasicek transition density is Gaussian in closed form (Vasicek, 1977; Yoshida, 1992; Aït-Sahalia, 2002). For $\Delta = \Delta t$,

$$r_{i+1} \mid r_i \sim N(m_i^{\mathbb{Q}}, v^{\mathbb{Q}}), \tag{17}$$

where

$$m_i^{\mathbb{Q}} = \theta^{\mathbb{Q}} + (r_i - \theta^{\mathbb{Q}})e^{-\kappa^{\mathbb{Q}}\Delta}, \tag{18}$$

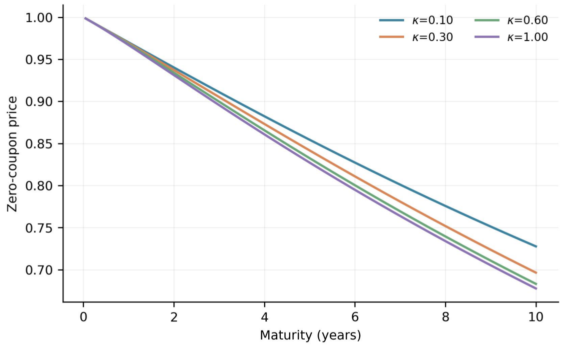

Fig. 5: Sensitivity of zero-coupon bond prices to mean-reversion speed κ.

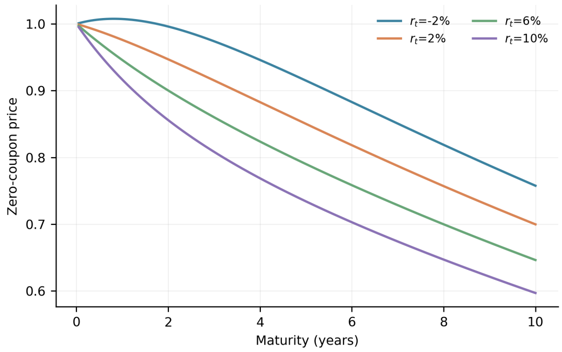

Fig. 6: Sensitivity of zero-coupon bond prices to the current short rate r_t.

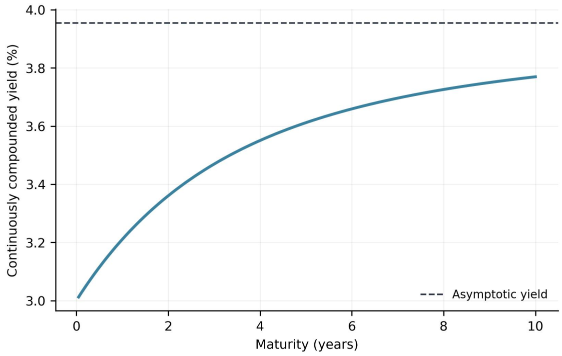

Fig. 7: Sample Vasicek term structure at a fixed observation time. The dashed line is the asymptotic yield.

and

$$v^{\mathbb{Q}} = \frac{\sigma^2}{2\kappa^{\mathbb{Q}}}\left(1 - e^{-2\kappa^{\mathbb{Q}}\Delta}\right). \tag{19}$$

Conditioning on r_0, the exact log-likelihood is

$$\begin{aligned}\ell_{\mathrm{EX}}(\phi^{\mathbb{Q}} \mid r) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln v^{\mathbb{Q}} \\ &- \frac{1}{2v^{\mathbb{Q}}}\sum_{i=0}^{n-1}\left(r_{i+1} - m_i^{\mathbb{Q}}\right)^2.\end{aligned} \tag{20}$$

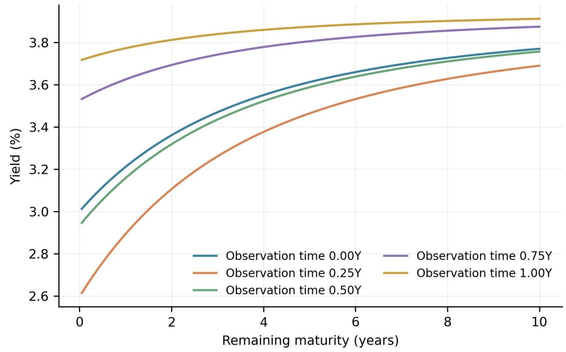

Fig. 8: Evolution of the Vasicek term structure as the observed short rate changes along a simulated path.

This objective is maximized numerically over $\kappa^{\mathbb{Q}} > 0$ and $\sigma > 0$. Since the exact transition law is used, EMLE does not incur time-discretization error when the data-generating process is correctly specified.

3.2 Euler-Maruyama Maximum Likelihood Estimation

EMMLE replaces the exact transition with the Euler-Maruyama discretization (Kloeden & Platen, 1992):

$$\begin{aligned}r_{i+1} = r_i + \kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - r_i)\Delta \\ + \sigma\sqrt{\Delta}\varepsilon_i, \qquad \varepsilon_i \sim N(0,1).\end{aligned} \tag{21}$$

The corresponding Gaussian pseudo-log-likelihood is

$$\begin{aligned}\ell_{\mathrm{EM}}(\phi^{\mathbb{Q}} \mid r) = &-\frac{n}{2}\ln(2\pi\sigma^2\Delta) \\ &- \frac{1}{2\sigma^2\Delta}\sum_{i=0}^{n-1}\begin{bmatrix} r_{i+1} - r_i \\ -\kappa^{\mathbb{Q}}(\theta^{\mathbb{Q}} - r_i)\Delta \end{bmatrix}^2.\end{aligned} \tag{22}$$

The Euler objective is easy to evaluate, but its accuracy depends on the sampling interval because it is based on a first-order discretization (Kloeden & Platen, 1992; Aït-Sahalia, 2002).

3.3 Quadratic-Variation-Initialized Two-Stage Likelihood

QV-2S is a simplified two-stage baseline in which a quadratic-variation proxy supplies an initial volatility scale and the remaining parameters are refined by exact-likelihood optimization. It is illustrative rather than a state-of-the-art realized-volatility estimator because its first stage omits the bias corrections of more advanced procedures

(Phillips & Yu, 2009; Barndorff-Nielsen & Shephard, 2002).

Let $T = n\Delta$. The discrete quadratic variation is

$$QV_T = \sum_{i=0}^{n-1} (r_{i+1} - r_i)^2, \tag{23}$$

which gives

$$\widehat{\sigma}^2_{QV} = \frac{QV_T}{T}. \tag{24}$$

For sufficiently fine grids, QV_T approximates $\sigma^2 T$. At finite sampling frequencies, however, this quantity is only an initializer because mean reversion and coarse observation can distort the approximation. The second stage maximizes the same exact likelihood used by EMLE, initialized at $\widehat{\sigma}_{QV}$. Consequently, any numerical difference between EMLE and QV-2S is an initialization and local-optimization effect rather than a difference in the statistical objective.

4. EXPERIMENTAL DESIGN AND RESULTS

4.1 Synthetic Data Simulation

The synthetic benchmark contains 1.2 million independent Vasicek paths simulated under $\mathbb{Q}$ using the exact Gaussian transition on a daily grid with $\Delta t = 1/252$ over one year. Parameters are sampled independently from

$$\theta^{\mathbb{Q}} \sim U[-1.0, 1.0], \qquad \sigma \sim U[0.005, 0.03], \tag{25}$$

$$\kappa^{\mathbb{Q}} \sim U[0.05, 1.00], \qquad r_0 \sim U[-0.05, 0.10]. \tag{26}$$

The corpus is split into 80% training and 20% validation paths. A further 10,000 independently simulated paths from the same design are reserved for testing. These ranges constitute a deliberately broad stress-test domain rather than a realistic market prior. The results therefore measure in-distribution synthetic recovery and do not establish robustness to distribution shift or empirical realism.

4.2 KAN and MLP Architecture and Inputs

KANs replace fixed scalar edge weights with learnable univariate spline functions (Liu et al., 2024; Bodner et al., 2024). The KAN has two hidden layers of widths 32 and 16, cubic B-spline edge functions, grid size 8, and a three-node output layer. Each path is summarized by 12 features: sample mean, sample standard deviation, lag-1 autocorrelation, minimum, maximum, initial level, terminal level, mean increment, standard deviation of increments, skewness, excess kurtosis, and an autocorrelation-based half-life proxy.

Input features are standardized using the training-set mean and standard deviation only; the fitted scaling is applied unchanged to validation and test observations. The transformed learning target is

$$z = (\theta^{\mathbb{Q}}, \log\sigma, \log\kappa^{\mathbb{Q}}), \tag{27}$$

and the predictions are mapped back through

$$\widehat{\sigma} = \exp(\widehat{\log\sigma}), \qquad \widehat{\kappa}^{\mathbb{Q}} = \exp(\widehat{\log\kappa^{\mathbb{Q}}}). \tag{28}$$

No clipping based on the true test-domain parameters is applied.

For scaled targets $\widetilde{z}_i$ and network predictions $\widehat{\widetilde{z}}_i$, the training loss is

$$\mathscr{L} = \frac{1}{N}\sum_{i=1}^{N} \left\| \widehat{\widetilde{z}}_i - \widetilde{z}_i \right\|^2. \tag{29}$$

The KAN is trained with Adam, learning rate 10^{-3}, batch size 512, weight decay 10^{-6}, and 50 epochs. The master random seed is 42.

The MLP uses the same 12 features, split, standardization, transformed targets, loss, optimizer, learning rate, batch size, weight decay, and number of epochs. It has hidden widths 32 and 16 with ReLU activations and a three-node output layer. Equal widths do not imply equal capacity: the spline coefficients on every KAN edge result in 11,379 trainable parameters, whereas the MLP has 995. Thus, the MLP is a matched-width supervised baseline rather than a parameter-count-matched baseline.

4.3 Numerical Optimization Details

For the likelihood estimators, the negative log-likelihood is minimized on the transformed scale

$$\psi = (\theta, \eta_\sigma, \eta_\kappa), \qquad \eta_\sigma = \log\sigma, \quad \eta_\kappa = \log\kappa. \tag{30}$$

The initial value for θ is the sample mean. The initial volatility is

$$\sigma_0 = \frac{\mathrm{sd}(r_{i+1} - r_i)}{\sqrt{\Delta}}, \tag{31}$$

except for QV-2S, which uses the quadratic-variation initializer. When the lag-1 sample auto-

Table. 1: Convergence and admissibility diagnostics

Method	Paths	Conv.	Admiss.	Med. iter.
EMLE	10,000	1.000	1.000	15
EMMLE	10,000	1.000	1.000	15
QV-2S	10,000	1.000	1.000	16
KAN	10,000	1.000	1.000	–
MLP	10,000	1.000	1.000	–

Note: Median iterations apply only to the likelihood estimators. KAN and MLP were trained for 50 epochs.

correlation satisfies $0 < \widehat{\rho}_1 < 1$, the initial mean-reversion speed is

$$\kappa_0 = -\frac{\log \widehat{\rho}_1}{\Delta}; \tag{32}$$

otherwise, it is set to the midpoint of the simulation interval.

Optimization uses L-BFGS-B with at most 1,000 iterations and tolerance 10^{-8}. To maintain comparability with the supervised estimators trained on the synthetic domain, the likelihood estimates are bounded by

$$\begin{aligned} &\theta \in [-1,1], \quad \sigma \in [0.005, 0.03], \\ &\kappa \in [0.05, 1.00]. \end{aligned} \tag{33}$$

No test information is used in estimation. All five methods are evaluated on the same 10,000 paths.

4.4 Convergence and Evaluation Completeness

Table 1 reports complete numerical coverage. Every method returns an admissible estimate for every test path. The likelihood procedures report 100% convergence, with median iteration counts of 15–16. KAN and MLP are both evaluated after the specified 50 training epochs.

4.5 Parameter-Recovery Results

Table 2 and Fig. 9 report parameter RMSE. KAN obtains the lowest RMSE for all three parameters: 0.1653 for θ, 0.00081 for σ, and 0.2058 for κ. The MLP is nearly indistinguishable for θ and κ, with RMSE values of 0.1653 and 0.2060, respectively. Relative to the best-performing likelihood benchmark for each parameter, KAN reduces RMSE by approximately 33.0% for θ, 3.0% for σ, and 43.6% for κ. The volatility gain is therefore small, while the improvements for the long-run mean and mean-reversion speed are economically and numerically larger.

The MLP results qualify an architecture-specific interpretation. Supervised learning from

Table. 2: Parameter RMSE on 10,000 synthetic test paths

Method	θ	σ	κ
EMLE	0.2465	0.00089	0.3653
EMMLE	0.2466	0.00088	0.3659
QV-2S	0.2468	0.00084	0.3646
KAN	**0.1653**	**0.00081**	**0.2058**
MLP	0.1653	0.00084	0.2060

the synthetic prior accounts for most of the improvement, and our experiment provides little overall parameter-RMSE separation between KAN and MLP. The KAN biases are 0.0150 for θ, 6.85×10^{-5} for σ, and -0.0483 for κ; the corresponding MLP biases are 0.0021, 3.47×10^{-5}, and -0.0442. Thus, lower RMSE does not imply absence of shrinkage or systematic error.

4.6 Mean-Reversion Regime Analysis

Table 3 decomposes errors by the true κ regime. In the low-κ regime, KAN has the lowest RMSE for all three parameters, including a κ RMSE of 0.1496 compared with 0.5924 for EMLE. In the medium regime, MLP is slightly better for θ and κ, while KAN is best for σ. In the high-κ regime, MLP has the lowest θ RMSE, KAN retains the lowest σ RMSE, but QV-2S has the lowest κ RMSE. KAN's high-regime κ RMSE of 0.2670 is approximately 26.6% higher than QV-2S's 0.2109. The supervised advantage is therefore strongest when slow mean reversion makes θ and κ difficult to identify from a one-year path; it does not extend uniformly to high-κ recovery.

4.7 Synthetic Price and Yield Error Evaluation

For each test path j, the terminal simulated short rate $r_{j,n}$ is treated as the current state. True and estimated prices and yields are computed on

$$\mathscr{T} = \{0.25, 0.5, 1, 2, 5, 10\} \text{ years.} \tag{34}$$

For method m and maturity τ_ℓ, the comparison is between

$$P^{\text{true}}_{j,\ell} = P(\tau_\ell; r_{j,n}, \phi^{\mathbb{Q}}_j) \tag{35}$$

and

$$\widehat{P}^{(m)}_{j,\ell} = P(\tau_\ell; r_{j,n}, \widehat{\phi}^{\mathbb{Q}}_{j,m}), \tag{36}$$

with an analogous comparison for yields. Yield errors are converted to basis points by multiplying decimal yield differences by 10,000.

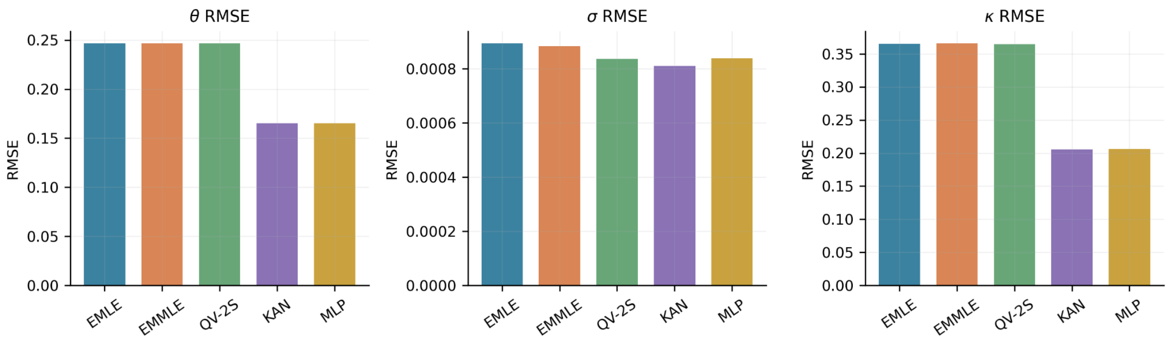

Fig. 9: Parameter RMSE across likelihood and supervised recovery methods. The scale differs across parameter panels.

Table. 3: Parameter RMSE by true mean-reversion regime

Regime	Method	Paths	θ	σ	κ
Low κ	EMLE	2,068	0.3661	0.00080	0.5924
	EMMLE	2,068	0.3674	0.00080	0.5935
	QV-2S	2,068	0.3684	0.00080	0.5915
	KAN	2,068	**0.2663**	**0.00078**	**0.1496**
	MLP	2,068	0.2724	0.00081	0.1518
Medium κ	EMLE	3,665	0.2476	0.00086	0.3375
	EMMLE	3,665	0.2468	0.00085	0.3368
	QV-2S	3,665	0.2485	0.00083	0.3370
	KAN	3,665	0.1367	**0.00081**	0.1413
	MLP	3,665	**0.1326**	0.00083	**0.1396**
High κ	EMLE	4,267	0.1575	0.00096	0.2116
	EMMLE	4,267	0.1576	0.00095	0.2134
	QV-2S	4,267	0.1545	0.00085	**0.2109**
	KAN	4,267	0.1166	**0.00082**	0.2670
	MLP	4,267	**0.1141**	0.00086	0.2674

Note: Low, medium, and high regimes are $\kappa \in [0.05, 0.25]$, $\kappa \in (0.25, 0.60]$, and $\kappa \in (0.60, 1.00]$, respectively.

Table 4 shows that QV-2S has the lowest yield RMSE at 0.25 years. KAN is lowest from 0.5 through 2 years, while MLP is lowest at 5 and 10 years. The separation from likelihood methods increases with maturity, but KAN does not dominate the MLP at the long end. Table 5 shows a related but not identical pattern: QV-2S is marginally best at 0.25 years, after which KAN has the lowest price RMSE at every reported maturity.

Table 6 summarizes all maturities jointly. KAN has the lowest aggregate price RMSE at 227.70, a reduction of approximately 30.2% relative to QV-2S and 11.5% relative to MLP. MLP has the lowest aggregate yield RMSE at 403.18 basis points, compared with 405.79 for KAN. The two supervised methods have almost identical yield MAE, but their bias patterns differ: KAN has a positive aggregate yield bias of 36.62 basis points, whereas MLP has a bias of -8.49 basis points. KAN also has a negative aggregate price bias of -11.92, while MLP's aggregate price bias is -0.46.

4.8 Paired Bootstrap Comparisons

The common test paths permit paired inference. Table 7 reports path-clustered percentile intervals from 2,000 bootstrap repetitions for $\mathrm{RMSE}(m) - \mathrm{RMSE}(\mathrm{KAN})$. Positive price differences favor KAN. All four price intervals are strictly positive, including the MLP-minus-KAN difference of 29.51 with a 95% interval of $[12.65, 45.63]$. KAN's price-RMSE advantage is therefore robust to paired resampling of the common test paths.

For yields, KAN is significantly better than each likelihood benchmark by approximately 101–103 basis points of aggregate RMSE. The MLP-minus-KAN difference is -2.60 basis points with a 95% interval of $[-4.37, -0.84]$, confirming that MLP has a small but statisti-

Table. 4: Yield RMSE in basis points across maturities

Method	0.25Y	0.5Y	1Y	2Y	5Y	10Y
EMLE	32.83	66.35	134.67	270.99	626.29	1030.15
EMMLE	32.76	66.22	134.47	270.85	626.83	1031.49
QV-2S	**32.57**	65.87	133.83	269.57	623.55	1026.55
KAN	32.92	**65.13**	**127.49**	**243.45**	518.00	799.26
MLP	33.71	66.37	128.98	244.15	**514.87**	**792.76**

Table. 5: Zero-coupon price RMSE across maturities

Method	0.25Y	0.5Y	1Y	2Y	5Y	10Y
EMLE	0.000825	0.003398	0.015084	0.088931	3.766605	801.044132
EMMLE	0.000823	0.003390	0.015041	0.088520	3.755151	801.541674
QV-2S	**0.000818**	0.003370	0.014946	0.087868	3.724157	798.689721
KAN	0.000820	**0.003272**	**0.013628**	**0.072696**	**2.809346**	**557.740314**
MLP	0.000849	0.003406	0.014379	0.078089	3.035413	630.012637

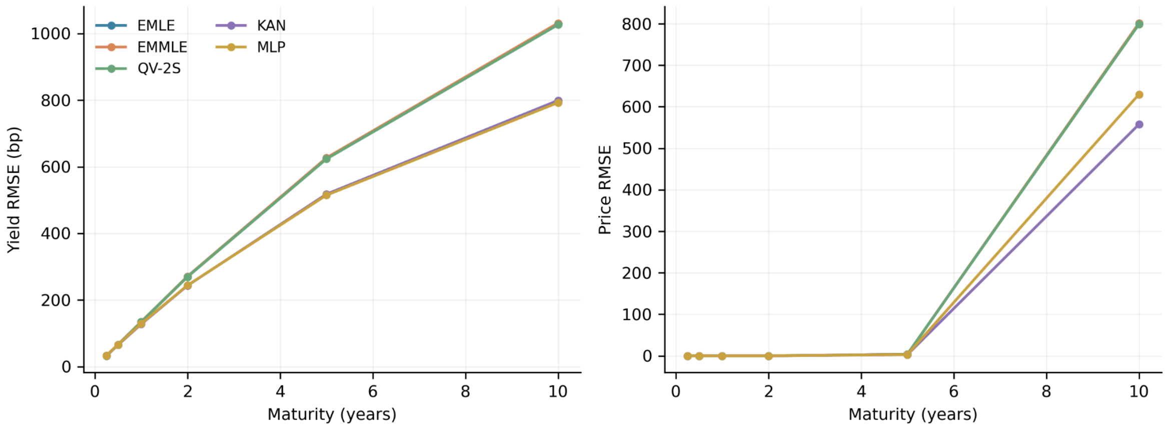

Fig. 10: Yield RMSE in basis points and zero-coupon price RMSE by maturity. Long-maturity absolute price errors are large because the broad synthetic parameter domain permits extreme model-implied bond prices.

Table. 6: Aggregate downstream errors across all paths and maturities

	Price error			Yield error (bp)		
Method	RMSE	MAE	Bias	RMSE	MAE	Bias
EMLE	327.0285	37.5018	10.5929	508.35	266.53	0.09
EMMLE	327.2316	37.3800	9.9126	508.88	267.15	1.11
QV-2S	326.0673	37.1264	10.3272	506.39	265.80	0.85
KAN	**227.6994**	**25.8431**	-11.9176	405.79	234.58	36.62
MLP	257.2046	28.4366	**-0.4591**	**403.18**	**234.55**	-8.49

cally resolved yield-RMSE advantage. The bootstrap evidence therefore reinforces the metric-dependent conclusion rather than supporting uniform KAN dominance.

4.9 Pricing Interpretation and Limitations

The reported prices and yields are synthetic downstream diagnostics, not empirical term-structure validation. The comparison is against the true Vasicek prices and yields generated from known risk-neutral parameters. It therefore isolates the propagation of recovery error under correct model specification, but it does not evaluate quoted bonds, observed yield curves, liquidity effects, measurement noise, or misspecification.

The broad $\theta^{\mathbb{Q}} \sim U[-1,1]$ domain is consequential. It admits negative long-end yields and very large long-maturity prices, so absolute price RMSE at 10 years is dominated by extreme price levels. Yield RMSE, maturity-specific errors,

Table. 7: Paired path-level bootstrap comparisons against KAN

Metric	Comparator minus KAN	Difference	95% interval
Price RMSE	EMLE	99.33	[82.10, 116.54]
Price RMSE	EMMLE	99.53	[81.31, 117.02]
Price RMSE	QV-2S	98.37	[80.34, 116.22]
Price RMSE	MLP	29.51	[12.65, 45.63]
Yield RMSE (bp)	EMLE	102.56	[95.47, 109.33]
Yield RMSE (bp)	EMMLE	103.10	[96.42, 110.24]
Yield RMSE (bp)	QV-2S	100.60	[93.57, 107.38]
Yield RMSE (bp)	MLP	-2.60	[-4.37, -0.84]

Note: Intervals use 2,000 bootstrap repetitions with the test path as the resampling unit. Positive values indicate larger RMSE than KAN.

MAE, and bias should therefore be read alongside aggregate price RMSE. The reported price values must not be interpreted as errors for a conventional unit-price bond market without conditioning on a realistic parameter domain.

The supervised estimators also learn the simulation prior. Their lower RMSE for weakly identified parameters can reflect amortization and shrinkage rather than universally more informative estimation. The near equality of KAN and MLP parameter RMSE, the MLP advantage for aggregate yield RMSE, the likelihood advantage for high-regime κ, and the unequal network parameter counts all prevent a strong architecture-specific conclusion. A parameter-count-matched MLP, out-of-distribution parameter tests, alternative path lengths, noisy or irregular observations, and direct market calibration are required before claiming a general KAN advantage.

Finally, QV-2S and EMLE optimize the same exact likelihood after different initialization. Their close results should be interpreted as a numerical robustness check, not as evidence for a distinct asymptotic estimator. A genuinely different two-stage method would require a separate second-stage rule.

5. CONCLUSION

This paper examined whether a spline-based KAN can recover risk-neutral Vasicek parameters from engineered features of synthetic short-rate paths. On 10,000 in-distribution test paths, KAN achieved the lowest overall RMSE for θ, σ, and κ, with the largest gains over likelihood methods for the long-run mean and mean-reversion speed. The matched-width MLP produced almost identical overall parameter accuracy, indicating that most of the gain is attributable to supervised amortized recovery rather than clearly to the KAN architecture alone.

Downstream performance was metric-dependent. KAN delivered the lowest aggregate zero-coupon price RMSE and significantly outperformed all comparators in paired bootstrap comparisons. The MLP, however, delivered a slightly but significantly lower aggregate yield RMSE, while likelihood estimation remained competitive at the shortest maturity and for high-regime κ. The defensible conclusion is therefore that supervised inverse-calibration mappings can improve synthetic recovery and downstream pricing under the chosen prior, but KAN does not uniformly dominate a conventional neural network or likelihood estimation.

Future work should use realistic risk-neutral parameter domains, parameter-count-matched architectures, multiple path lengths and sampling frequencies, measurement noise, out-of-distribution tests, hybrid KAN-likelihood refinement, and direct calibration to observed term structures. These extensions are necessary to determine whether the synthetic gains translate into reliable empirical fixed-income applications.

6. ACKNOWLEDGMENT

The authors acknowledge the use of an AI-based writing assistant for language polishing and structural editing. The authors take full responsibility for the analysis, interpretation, and conclusions.

7. REFERENCES

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, *70*(1), 223–262.

Barndorff-Nielsen, O. E., & Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*, *64*(2), 253–280.

Bodner, A. D., Tepsich, A. S., Spolski, J. N., & Pourteau, S. (2024). Convolutional Kolmogorov-Arnold Networks. *arXiv:2406.13155*.

Brigo, D., & Mercurio, F. (2006). *Interest rate models–Theory and practice* (2nd ed.). Springer, Berlin, Germany.

Duffie, D., & Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, *6*(4), 379–406.

Kloeden, P. E., & Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer, Berlin, Germany.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., & Tegmark, M. (2024). KAN: Kolmogorov-Arnold Networks. *arXiv:2404.19756*.

Phillips, P. C. B., & Yu, J. (2009). A two-stage realized volatility approach to estimation of diffusion processes from discrete observations. *Journal of Econometrics*, *150*(2), 139–150.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, *5*(2), 177–188.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observations. *Journal of Multivariate Analysis*, *41*(2), 220–242.