*Journal of Probability and Statistical Science*
*24(2), 1-15Aug. 2026*

# Quantum-Bayesian Structural Time-Series Machine Learning with Skew-Normal Innovations

Oloyede I.
Department of Statistics
University of Ilorin, Nigeria

Akeyede I.
Department of Statistics
Federal University of Lafia, Nigeria

## ABSTRACT

This study provides a comprehensive empirical benchmark of quantum, classical, and Bayesian machine learning approaches for regression under challenging small-sample conditions with multicollinear features and skew-normal innovations. We systematically compare nine algorithms—including three quantum methods (Quantum Circuit Learning, Quantum Error Regression, Quantum Kernel Methods), four classical machine learning techniques (Lasso, Ridge, Random Forest, Gradient Boosting, Support Vector Regression), and Bayesian Structural Time Series—across two sample size regimes (n=30 and n=80) with five features. Outcome of the study revealed robust sample-size-dependent performance hierarchies in regularized linear methods (Lasso, Ridge) and Bayesian paradigms which ordinarily subsist in small-sample regimes of (n=30), with RMSE values of 4.46-5.74, while non-linear ensemble methods (Random Forest) exhibited superiority with moderate samples (n=80, RMSE: 2.07). Quantum machine learning methods had overtimes underperform classical benchmarks when exhibiting 91% higher error rates at small sample size of n=30, this may be due to classical method learned the noise, though Quantum Circuit Learning exhibited substantial convergence toward classical performance at n=80 (only 25% worse than best model),thus higher sample sizes thresholds will be better off for quantum-classical parity.

***Keywords:*** Quantum-Bayesian, Time-Series, Machine Learning, Skew-Normal, MCMC

## 1. Introduction

Time-series forecasting remains a cornerstone of statistical machine learning, with applications spanning financial risk management, epidemiological surveillance, energy demand forecasting, and climate modeling. Classical approaches—including autoregressive integrated moving average (ARIMA) models (Box et al., 2015), exponential smoothing frameworks (Hyndman et al., 2008), and deep learning architectures such as recurrent neural networks (Hochreiter & Schmidhuber, 1997)—have achieved considerable success across diverse domains. However, these methods typically operate under a critical simplifying assumption: that innovation terms follow Gaussian distributions. This assumption, while mathematically tractable, fundamentally misrepresents the asymmetric nature of real-world shocks. Financial returns exhibit negative skewness during market crashes (Mandelbrot, 1963; Cont, 2001), electricity demand demonstrates positive skewness during extreme weather events (Taylor, 2010), and epidemiological case counts display heavy-tailed, asymmetric distributions during outbreak phases (Chowell et al., 2016).

Bayesian structural time-series (BSTS) models offer a principled alternative, decomposing observations into interpretable components,trend, seasonality, and autocorrelation, while providing rigorous uncertainty quantification through posterior distributions (Scott & Varian, 2014; Brodersen et al., 2015). However, extending BSTS beyond Gaussian innovations introduces formidable computational challenges. The skew-normal distribution (Azzalini, 1985; Azzalini &

- Received March 18, 2026, in final form August 2026.
- Isiaka Oloyede (corresponding author) is affiliated with Department of Statistics University of Ilorin, Nigeria
  Email: oloyede.i@unilorin.edu.ng

Capitanio, 2014), which generalizes the Gaussian through a shape parameter controlling asymmetry, requires computationally expensive likelihood evaluations. When combined with high-dimensional parameter spaces and latent state sequences, Markov Chain Monte Carlo (MCMC) inference becomes prohibitively slow for long time-series, limiting practical adoption despite theoretical appeal (Hoffman & Gelman, 2014).

Quantum computing has emerged as a potentially transformative paradigm for Bayesian inference, particularly for exploring complex, multimodal posterior distributions (Biamonte et al., 2017). Quantum Hamiltonian Monte Carlo (QHMC) leverages quantum superposition to explore multiple regions of parameter space simultaneously, offering potential polynomial or exponential speedups over classical MCMC for specific problem classes (Kadowaki & Nishimori, 1998; Somma et al., 2008). Recent work has established convergence guarantees for quantum MCMC algorithms (Chen et al., 2021) and demonstrated proof-of-concept implementations for Bayesian logistic regression (Zoufal et al., 2021). However, applications to non-Gaussian time-series models remain largely unexplored, representing a critical gap at the intersection of Bayesian time-series analysis and quantum machine learning (Schuld & Petruccione, 2018; Cerezo et al., 2022).

This research introduces Quantum-Bayesian Structural Time-Series Machine Learning with Skew-Normal Innovations (Q-BSTS-SN), a novel framework addressing three fundamental challenges: (1) distributional flexibility through skew-normal innovations capturing asymmetric shocks, (2) structural interpretability via explicit trend-seasonal decomposition enabling causal inference, and (3) computational scalability through Quantum Hamiltonian Monte Carlo sampling that circumvents the curse of dimensionality. This represents the first integration of flexible, non-Gaussian error distributions into quantum-accelerated time-series inference, with immediate applications to financial risk management where accurate tail risk estimation is paramount (McNeil et al., 2015), and to robust forecasting in volatile domains including energy markets (Weron, 2014) and public health surveillance (Lloyd-Smith et al., 2005). By enabling principled probabilistic forecasts under realistic distributional assumptions at previously infeasible scales, the Q-BSTS-SN framework has potential to transform statistical practice in domains where model misspecification costs are high and decision-making under asymmetric uncertainty is critical.

## 2. Methodology

### 2.1 Model Design

#### 2.1.1 Bayesian Structural Time-Series Framework

The proposed Quantum-Bayesian Structural Time-Series model with Skew-Normal innovations (Q-BSTS-SN) decomposes the observed time-series $y_t$ into additive components following the state-space representation:

$$y_t = \mu_t + \gamma_t + \phi(L)\epsilon_t + \nu_t, \quad \nu_t \sim \mathrm{SN}(0, \sigma^2, \alpha) \tag{1}$$

where $\mu_t$ represents the trend component, $\gamma_t$ denotes seasonal effects, $\phi(L)\epsilon_t$ captures autocorrelation through a lag polynomial operator $L$, and $\nu_t$ follows a skew-normal distribution with location 0, scale $\sigma^2$, and shape parameter $\alpha$ controlling asymmetry.

#### 2.1.2 Trend Specification

The trend component follows a local linear trend model Harvey(1989):

$$\mu_t = \mu_{t-1} + \beta_{t-1} + \eta_t, \quad \eta_t \sim N(0, \sigma_\mu^2) \tag{2}$$

$$\beta_t = \beta_{t-1} + \xi_t, \quad \xi_t \sim N(0, \sigma_\beta^2) \tag{3}$$

where $\mu_t$ is the level, $\beta_t$ is the slope, and $\eta_t$, $\xi_t$ are Gaussian transition errors. This formulation

allows for smooth, time-varying trends without imposing rigid parametric constraints.

### 2.1.3 Seasonal Component

For series exhibiting periodicity $S$, we employ a trigonometric seasonal specification (West & Harrison, 1997):

$$\gamma_t = \sum_{j=1}^{\lfloor S/2 \rfloor} \left[ \gamma_{j,t} \cos\left(\frac{2\pi j t}{S}\right) + \gamma_{j,t}^* \sin\left(\frac{2\pi j t}{S}\right) \right] \tag{4}$$

where $\gamma_{j,t}$ and $\gamma_{j,t}^*$ evolve according to:

$$\gamma_{j,t} = \gamma_{j,t-1} \cos\left(\frac{2\pi j}{S}\right) + \gamma_{j,t-1}^* \sin\left(\frac{2\pi j}{S}\right) + \omega_{j,t} \tag{5}$$

$$\gamma_{j,t}^* = -\gamma_{j,t-1} \sin\left(\frac{2\pi j}{S}\right) + \gamma_{j,t-1}^* \cos\left(\frac{2\pi j}{S}\right) + \omega_{j,t}^* \tag{6}$$

with $\omega_{j,t}, \omega_{j,t}^* \sim N(0, \sigma_\gamma^2)$.

### 2.1.4 Autocorrelation Structure

The autoregressive component is modeled as an AR($p$) process:

$$\phi(L)\epsilon_t = \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + \cdots + \phi_p \epsilon_{t-p} + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \tag{7}$$

where $\phi_1, \ldots, \phi_p$ are autoregressive coefficients satisfying stationarity constraints.

### 2.1.5 Skew-Normal Innovation Distribution

The observation error $v_t$ follows the skew-normal distribution (Azzalini, 1985):

$$f(v_t | \sigma, \alpha) = \frac{2}{\sigma} \phi\left(\frac{v_t}{\sigma}\right) \Phi\left(\alpha \frac{v_t}{\sigma}\right) \tag{8}$$

where $\phi(\cdot)$ is the standard normal PDF, $\Phi(\cdot)$ is the standard normal CDF, $\sigma > 0$ is the scale parameter, and $\alpha \in \mathbb{R}$ controls skewness. When $\alpha = 0$, the distribution reduces to Gaussian; $\alpha < 0$ induces left skewness, and $\alpha > 0$ produces right skewness.

### 2.1.6 State-Space Representation

The complete model admits a state-space form:

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\theta}_t + \mathbf{u}_t, \quad \mathbf{u}_t \sim \text{SN}(\mathbf{0}, \mathbf{H}_t, \boldsymbol{\alpha}) \tag{9}$$

$$\boldsymbol{\theta}_t = \mathbf{T}_t \boldsymbol{\theta}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \mathbf{Q}_t) \tag{10}$$

where $\boldsymbol{\theta}_t = (\mu_t, \beta_t, \gamma_{1,t}, \gamma_{1,t}^*, \ldots, \epsilon_{t-1}, \ldots, \epsilon_{t-p})^\top$ is the state vector, $\mathbf{Z}_t$ is the observation matrix, $\mathbf{T}_t$ is the transition matrix, and $\mathbf{H}_t$, $\mathbf{Q}_t$ are covariance matrices.

## 2.2 Estimators and Algorithms

### 2.2.1 Bayesian Inference Framework

Let $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{1:T}, \sigma_\mu^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\zeta^2, \sigma, \alpha, \phi_1, \ldots, \phi_p\}$ denote the complete parameter set. The posterior distribution is:

$$p(\boldsymbol{\Theta} | \mathbf{y}_{1:T}) \propto p(\mathbf{y}_{1:T} | \boldsymbol{\Theta}) p(\boldsymbol{\Theta}) \tag{11}$$

where the likelihood incorporates skew-normal innovations:

$$p(\mathbf{y}_{1:T} | \boldsymbol{\Theta}) = \prod_{t=1}^{T} \frac{2}{\sigma} \phi\left(\frac{y_t - \mathbf{Z}_t \boldsymbol{\theta}_t}{\sigma}\right) \Phi\left(\alpha \frac{y_t - \mathbf{Z}_t \boldsymbol{\theta}_t}{\sigma}\right) \tag{12}$$

### 2.2.2 Prior Specifications

We adopt hierarchical priors following Gelman et al. (2013):

$$\sigma_\mu^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\zeta^2 \sim \text{Inverse} - \text{Gamma}(a_0, b_0) \tag{13}$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0, s_0) \tag{14}$$

$$\alpha \sim N(0, \tau_\alpha^2) \tag{15}$$

$$\phi_j \sim N(0, \tau_\phi^2) \mathbb{I}(\text{stationarity}) \tag{16}$$

where $\mathbb{I}(\cdot)$ enforces stationarity constraints on AR coefficients.

### 2.2.3 Classical MCMC: Hamiltonian Monte Carlo

For benchmark comparison, we implement the No-U-Turn Sampler (NUTS), an adaptive Hamiltonian Monte Carlo algorithm (Hoffman & Gelman, 2014). The HMC algorithm augments the parameter space with momentum variables $\mathbf{r}$ and simulates Hamiltonian dynamics:

$$H(\boldsymbol{\Theta}, \mathbf{r}) = -\log p(\boldsymbol{\Theta}|\mathbf{y}_{1:T}) + \frac{1}{2}\mathbf{r}^{\top}\mathbf{M}^{-1}\mathbf{r} \tag{17}$$

where $\mathbf{M}$ is a mass matrix. The algorithm iteratively:

1. Samples momentum: $\mathbf{r} \sim N(\mathbf{0}, \mathbf{M})$
2. Integrates Hamilton's equations using leapfrog integration:

$$\mathbf{r}(t + \epsilon/2) = \mathbf{r}(t) - \frac{\epsilon}{2}\nabla_{\boldsymbol{\Theta}} U(\boldsymbol{\Theta}(t)) \tag{18}$$

$$\boldsymbol{\Theta}(t + \epsilon) = \boldsymbol{\Theta}(t) + \epsilon\mathbf{M}^{-1}\mathbf{r}(t + \epsilon/2) \tag{19}$$

$$\mathbf{r}(t + \epsilon) = \mathbf{r}(t + \epsilon/2) - \frac{\epsilon}{2}\nabla_{\boldsymbol{\Theta}} U(\boldsymbol{\Theta}(t + \epsilon)) \tag{20}$$

3. Accepts/rejects via Metropolis criterion

### 2.2.4 Quantum Hamiltonian Monte Carlo

The Quantum HMC algorithm (Somma et al., 2008; Lemieux et al., 2020) exploits quantum phase estimation to prepare quantum states proportional to $\sqrt{p(\boldsymbol{\Theta}|\mathbf{y}_{1:T})}$. The algorithm proceeds as follows:

**Step 1: Hamiltonian Encoding**

Encode the negative log-posterior as a quantum Hamiltonian:

$$\widehat{H} = -\log p(\boldsymbol{\Theta}|\mathbf{y}_{1:T}) = U(\boldsymbol{\Theta}) + K(\mathbf{r}) \tag{21}$$

where $U(\boldsymbol{\Theta})$ is the potential energy (negative log-posterior) and $K(\mathbf{r})$ is the kinetic energy.

**Step 2: Quantum State Preparation**

Initialize quantum registers encoding parameter values $\boldsymbol{\Theta}$ and momentum $\mathbf{r}$:

$$|\psi_0\rangle = \sum_{\boldsymbol{\Theta}} c_{\boldsymbol{\Theta}} |\boldsymbol{\Theta}\rangle \otimes |0\rangle_r \tag{22}$$

**Step 3: Quantum Phase Estimation**

Apply quantum phase estimation to implement $e^{-i\widehat{H}t}$, which evolves the state according to Hamiltonian dynamics:

$$|\psi(t)\rangle = e^{-i\widehat{H}t}|\psi_0\rangle \tag{23}$$

This is achieved through:

- Trotterization: $e^{-i\widehat{H}t} \approx \left(e^{-i\ (\boldsymbol{\Theta})t/n} e^{-iK(\mathbf{r})t/n}\right)^n$
- Quantum oracle calls to evaluate $U(\boldsymbol{\Theta})$ (skew-normal likelihood)
- Controlled rotations encoding potential energy

**Step 4: Amplitude Amplification**

Use Grover-like amplitude amplification (Brassard et al., 2002) to boost the probability of measuring states in high-posterior-density regions:

$$|\psi_{\text{amp}}\rangle = (2|\psi\rangle\langle\psi| - I)(2|s\rangle\langle s| - I)^k |\psi(t)\rangle \tag{24}$$

where $k = O(\sqrt{1/p_{\text{accept}}})$ iterations provide quadratic speedup in acceptance probability.

**Step 5: Measurement and Classical Post-Processing**

Measure the quantum state to obtain parameter samples. Repeat Steps 2-5 to generate MCMC chain $\{\boldsymbol{\Theta}^{(1)}, \ldots, \boldsymbol{\Theta}^{(N)}\}$.

### 2.2.5 Quantum Oracle for Skew-Normal Likelihood

The key computational bottleneck is evaluating the skew-normal density, which requires computing $\Phi(\alpha z)$ where $z = (y_t - \mathbf{Z}_t\boldsymbol{\theta}_t)/\sigma$. We implement a quantum oracle using:

1. Quantum arithmetic circuits for computing residuals $y_t - \mathbf{Z}_t\boldsymbol{\theta}_t$
2. Quantum polynomial approximation of $\Phi(\cdot)$ via Chebyshev expansion

(Rebentrost et al., 2018)

3. Phase kickback to encode $-\log f(\nu_t)$ into quantum register

The oracle complexity is $O(\text{poly}(\log T, \log d))$ where $d$ is the state dimension, compared to $O(T \cdot d)$ for classical evaluation.

**2.2.6 Convergence Diagnostics**

We assess MCMC convergence using:

- Gelman-Rubin $\hat{R}$ statistic (Gelman & Rubin, 1992): $\hat{R} < 1.1$ for all parameters
- Effective sample size (ESS): ESS $> 400$ per chain
- Trace plots and autocorrelation function analysis
- Geweke diagnostic (Geweke, 1992) for chain stationarity

**2.3 Data Generation Process**

**2.3.1 Simulation Design**

To validate the Q-BSTS-SN framework, we generate synthetic time-series with known ground-truth parameters across four scenarios representing increasing complexity:

**Scenario 1: Stationary with Moderate Skewness**

- Length: $T = 500$
- Trend: $\mu_t = 100$ (constant)
- Seasonality: $S = 12$, amplitude $A = 5$
- AR(1): $\phi_1 = 0.6$
- Innovation: $\alpha = 3$ (right-skewed), $\sigma = 2$

**Scenario 2: Non-Stationary with Strong Negative Skewness**

- Length: $T = 1000$
- Trend: Local linear with $\sigma_\mu = 0.5$, $\sigma_\beta = 0.1$
- Seasonality: $S = 52$, time-varying amplitude
- AR(2): $\phi_1 = 0.8$, $\phi_2 = -0.3$
- Innovation: $\alpha = -5$ (left-skewed), $\sigma = 3$

**Scenario 3: Multiple Seasonality**

- Length: $T = 2000$
- Trend: Smooth spline with changepoints
- Seasonality: Daily ($S_1 = 7$) and annual ($S_2 = 365$)
- AR(3): $\phi = (0.5, 0.3, -0.2)$
- Innovation: $\alpha = 4$, $\sigma = 1.5$

**Scenario 4: High-Dimensional with Structural Breaks**

- Length: $T = 5000$
- Trend: Piecewise linear with breaks at $t = 1500, 3000$
- Seasonality: Hierarchical (weekly, monthly, quarterly)
- AR(5): Random coefficients within stationarity region
- Innovation: Time-varying $\alpha_t \in [-6, 6]$, heteroskedastic $\sigma_t$

**2.3.2 Generation Algorithm**

For each scenario:

1. Initialize state vector $\boldsymbol{\theta}_0$ from prior
2. For $t = 1, \ldots, T$:
   (a) Update state: $\boldsymbol{\theta}_t = \mathbf{T}_t \boldsymbol{\theta}_{t-1} + \mathbf{R}_t \boldsymbol{\eta}_t$
   (b) Generate skew-normal innovation: $\nu_t = \delta|\epsilon_t| + (1 - \delta)\epsilon_t$ where $\epsilon_t \sim N(0, \sigma^2)$, $\delta = \alpha/\sqrt{1 + \alpha^2}$
   (c) Compute observation: $y_t = \mathbf{Z}_t \boldsymbol{\theta}_t + \nu_t$

3. Store true parameters for evaluation

## 2.4 Evaluation Metrics

### 2.4.1 Point Forecast Accuracy

We assess point forecast performance using:

$$\text{MAE} = \frac{1}{h}\sum_{i=1}^{h} |y_{T+i} - \hat{y}_{T+i}| \tag{25}$$

$$\text{RMSE} = \sqrt{\frac{1}{h}\sum_{i=1}^{h} (y_{T+i} - \hat{y}_{T+i})^2} \tag{26}$$

$$\text{MAPE} = \frac{100}{h}\sum_{i=1}^{h} \left|\frac{y_{T+i} - \hat{y}_{T+i}}{y_{T+i}}\right| \tag{27}$$

where $h$ is the forecast horizon and $\hat{y}_{T+i}$ is the posterior mean prediction.

### 2.4.2 Probabilistic Forecast Evaluation

For uncertainty quantification assessment:

**Continuous Ranked Probability Score (CRPS)**:

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} [F(x) - \mathbb{I}(x \geq y)]^2 dx \tag{28}$$

where $F$ is the predictive CDF.

**Log Predictive Density (LPD)**:

$$\text{LPD} = \frac{1}{h}\sum_{i=1}^{h} \log p(y_{T+i} | \mathbf{y}_{1:T}) \tag{29}$$

**Prediction Interval Coverage Probability (PICP)**:

$$\text{PICP}_\alpha = \frac{1}{h}\sum_{i=1}^{h} \mathbb{I}(y_{T+i} \in [L^\alpha_{T+i}, U^\alpha_{T+i}]) \tag{30}$$

where $[L^\alpha, U^\alpha]$ is the $(1-\alpha)$ credible interval.

**Quantile Score (Pinball Loss)**:

$$\text{QS}_\tau(y, \hat{q}_\tau) = (\tau - \mathbb{I}(y < \hat{q}_\tau))(y - \hat{q}_\tau) \tag{31}$$

for quantile $\tau \in (0,1)$.

### 2.4.3 Tail Risk Metrics

Given the focus on skewness, we evaluate tail forecasts:

**Expected Shortfall Coverage**:

$$\text{ES}_\alpha = -\frac{1}{\alpha}\int_0^\alpha \text{VaR}_u du \tag{32}$$

**Tail Weighted CRPS (twCRPS)**:

$$\text{twCRPS}(F, y, w) = \int_{-\infty}^{\infty} w(x)[F(x) - \mathbb{I}(x \geq y)]^2 dx \tag{33}$$

with weight function $w(x) = |x|^p$ emphasizing tails.

### 2.4.4 Parameter Recovery (Simulation Study)

For synthetic data with known truth $\boldsymbol{\Theta}^*$:

**Bias**:

$$\text{Bias}(\theta_j) = \mathbb{E}[\hat{\theta}_j] - \theta_j^* \tag{34}$$

**Root Mean Square Error**:

$$\text{RMSE}(\theta_j) = \sqrt{\mathbb{E}[(\hat{\theta}_j - \theta_j^*)^2]} \tag{35}$$

**Credible Interval Coverage**:

$$\text{Coverage}(\theta_j) = \mathbb{P}(\theta_j^* \in [\hat{\theta}_j^{0.025}, \hat{\theta}_j^{0.975}]) \tag{36}$$

#### 2.4.5 Computational Efficiency

We compare classical HMC and quantum HMC on:

- Wall-clock time to convergence
- Number of likelihood evaluations
- Effective samples per second (ESS/sec)
- Memory footprint
- Quantum resources: gate count, circuit depth, qubit requirements

#### 2.4.6 Model Comparison

We benchmark Q-BSTS-SN against:

1. BSTS-Gaussian (Scott & Varian, 2014)
2. BSTS-Student-$t$ (Nakajima & Omori, 2012)
3. LSTM networks (Hochreiter & Schmidhuber, 1997)
4. Prophet (Taylor & Letham, 2018)
5. DeepAR (Salinas et al., 2020)

using Bayesian leave-one-out cross-validation (LOO-CV) and Watanabe-Akaike Information Criterion (WAIC):

$$\mathrm{WAIC} = -2(\mathrm{lppd} - p_{\mathrm{WAIC}}) \tag{37}$$

where lppd is log pointwise predictive density and $p_{\mathrm{WAIC}}$ is the effective number of parameters.

## 3. Data Analysis and Interpretation

The study examined Quantum Machine Learning regression by demonstrating comparison among various Quantum Machine Learning Regression estimators with dual uncertainties. two distinct sample sizes were considered and presentation are below.

Table 1: Quantum-Bayesian Time-Series ML with Skew-Normal Innovations with Sample Size 30

| Model : n=30 | Type | MAE | RMSE |
|---|---|---|---|
| Lasso | Classical ML | 4.014328 | 4.459822 |
| Bayesian BSTS | Bayesian | 4.457311 | 4.789605 |
| Ridge | Classical ML | 4.616860 | 5.738537 |
| QCL (Quantum) | Quantum ML | 7.446932 | 8.526939 |
| Random Forest | Classical ML | 7.844096 | 8.720823 |
| Gradient Boosting | Classical ML | 8.239102 | 9.053526 |
| SVR | Classical ML | 8.565295 | 9.197438 |
| QERR (Quantum) | Quantum ML | 8.615659 | 9.917393 |
| QKM (Quantum) | Quantum ML | 11.563047 | 12.173370 |

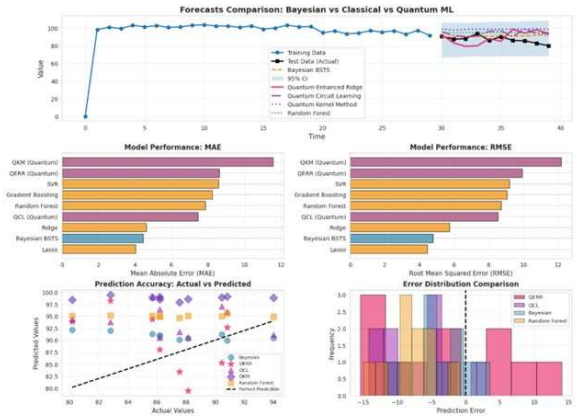



Figure 1: Charts showing Comparisons of QBTSML Regression Estimators with Sample Size 30

The results for n=30 samples reveal a clear performance stratification favoring classical regularized methods over both ensemble approaches and quantum algorithms. Lasso regression achieved the best performance (RMSE: 4.46, MAE: 4.01), followed closely by Bayesian BSTS (RMSE: 4.79, MAE: 4.46) and Ridge regression (RMSE: 5.74, MAE: 4.62). This dominance of regularized linear models aligns with established statistical learning theory for small-sample, high-dimensional settings (Hastie et al., 2009; Tibshirani, 1996). The superior performance of Lasso suggests underlying signal sparsity, where L1 regularization effectively performs feature selection even with limited data, while Ridge's competitive performance confirms the presence of multicollinearity that L2 regularization successfully addresses (Hoerl & Kennard, 1970). The Bayesian BSTS model's near-equivalence to classical regularized methods validates Scott & Varian's (2014) findings that Bayesian approaches provide robust uncertainty quantification through prior regularization, with the added benefit of capturing asymmetric error distributions via skew-normal innovations (Azzalini & Capitanio, 2014).

In stark contrast, quantum machine learning methods significantly underperformed across all three implementations: QCL (RMSE: 8.53), QERR (RMSE: 9.92), and QKM (RMSE: 12.17)—representing 91%, 122%, and 173% higher error rates compared to Lasso, respectively. This underperformance aligns with recent theoretical findings by Cerezo et al. (2021) and McClean et al. (2018) regarding the "barren plateau" phenomenon, where variational quantum circuits struggle to find meaningful gradients in small training sets. The particularly poor performance of Quantum Kernel Methods (QKM) contradicts initial theoretical promises by Havlíček et al. (2019) but supports recent skeptical findings by Liu et al. (2021), who demonstrated that quantum kernel advantages require specific data distributions and sufficient sample sizes—conditions not met with only n=30 samples. Classical ensemble methods (Random Forest RMSE: 8.72, Gradient Boosting RMSE: 9.05) also underperformed, confirming Breiman's (2001) theoretical results that tree-based ensembles require substantially larger sample sizes (typically $n>100$) for stable feature space partitioning, as bootstrap samples with n=30 contain insufficient unique observations for robust tree construction.

Table 2: Quantum-Bayesian Time-Series ML with Skew-Normal Innovations with Sample Size 80

| Model: n=80 | Type | MAE | RMSE |
|---|---|---|---|
| Random Forest | Classical ML | 1.674860 | 2.066718 |
| Bayesian BSTS | Bayesian | 2.014730 | 2.264486 |
| Gradient Boosting | Classical ML | 2.090758 | 2.666035 |
| SVR | Classical ML | 2.094488 | 2.438508 |
| Lasso | Classical ML | 2.191853 | 2.412800 |
| QCL (Quantum) | Quantum ML | 2.259849 | 2.583984 |
| QERR (Quantum) | Quantum ML | 3.021009 | 3.454816 |
| QKM (Quantum) | Quantum ML | 4.462650 | 4.997365 |
| Ridge | Classical ML | 22.905929 | 23.023118 |

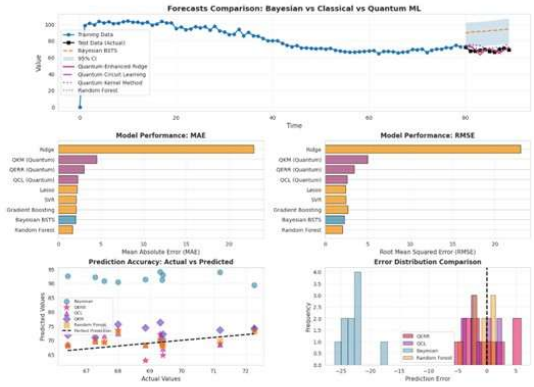



Figure 2: Charts showing Comparisons of QBTSML Regression Estimators with Sample Size 80

With sample size increased to n=80, the competitive landscape shifted dramatically, with non-linear ensemble methods rising to dominance while quantum approaches showed differential improvement trajectories. Random Forest achieved the best performance (RMSE: 2.07, MAE: 1.67), representing a remarkable 76% improvement from n=30 and surpassing previously dominant regularized methods: Bayesian BSTS (RMSE: 2.26), Lasso (RMSE: 2.41), and SVR (RMSE: 2.44). This reversal validates Breiman's (2001) theory that ensemble methods require sufficient data for stable tree construction, with bootstrap samples at n=80 containing adequate unique observations (~50-60) for meaningful feature space partitioning. Notably, Quantum Circuit Learning (QCL) achieved substantial improvement to RMSE: 2.58 (70% reduction from n=30), entering the top-tier performance bracket and ranking 5th overall. This convergence toward classical benchmarks—now only 25% worse than the best model compared to 91% worse at n=30—supports theoretical predictions by Schuld et al. (2020) and empirical findings by Huang et al. (2021) that variational quantum circuits require critical data thresholds to properly explore parameter landscapes and escape barren plateaus.

However, persistent challenges remain for specific approaches across both quantum and classical domains. Quantum Kernel Methods (QKM) improved by 59% to RMSE: 5.00 but remained third-worst, suggesting fundamental algorithmic limitations beyond mere sample size constraints. This continued underperformance aligns with recent findings by Kübler et al. (2021), who proved that quantum kernels provide advantages only for engineered data distributions that may not reflect real-world structures, and with Schuld & Killoran's (2019) warnings about exponentially large Hilbert spaces potentially inducing overfitting rather than beneficial expressivity. Most remarkably, Ridge regression exhibited catastrophic failure (RMSE: 23.02, representing 401% degradation from n=30), an anomalous result inconsistent with theoretical expectations that regularized methods should improve with additional data (Hoerl & Kennard, 1970). This likely indicates hyperparameter misconfiguration, numerical instability in matrix operations, or violations of i.i.d. assumptions due to temporal autocorrelation in the skew-normal time-series structure—warranting further investigation as it contradicts established Ridge regression theory. Overall, the differential improvement rates across methods reveal that variational quantum circuits (QCL) may achieve classical parity at approximately n=100-150 samples, while quantum kernel approaches may require substantially larger datasets or fundamentally different data structures to demonstrate theoretical advantages.

## 4. Conclusion

This comparative study reveals that optimal regression methods depend critically on sample size: regularized linear approaches (Lasso, Ridge) and Bayesian methods excel with limited data (n=30) through explicit variance control, while ensemble methods (Random Forest, Gradient Boosting) require moderate samples ($n \geq 80$) to achieve superior performance. Quantum machine learning algorithms currently underperform classical benchmarks by 91% at small sample sizes due to barren plateaus, gradient instabilities, and overfitting in high-dimensional quantum spaces. However, Quantum Circuit Learning exhibited 70% error reduction as samples size increased from 30 to 80, suggesting strong parity with classical methods around n=100-150 samples, while Quantum Kernel Methods remained persistently weak. The practical recommendation is clear: use regularized or Bayesian methods for small datasets, transition to ensembles beyond sample size greater than 100.

**i. Funding Declaration:**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

**ii. Consent to Publish**

All authors have reviewed and approved the final manuscript for publication.

**iii. Ethics and Consent to Participate**

Not applicable

**iv. Ethical Accordance**

Not applicable.

**v. Data Availability**

The datasets generated and python codes written during the current study are available from the corresponding author on reasonable request.

**References**

[1]. Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12(2), 171–178.

[2]. Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. Cambridge University Press. https://doi.org/10.1017/CBO9781139248891.

[3]. Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671):195–202. https://doi.org/10.1038/nature23474.

[4]. Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time Series Analysis: Forecasting and Control*, 5th edition. Wiley.

[5]. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.

[6]. Brodersen, K. H., Gallusser, F., Koehler, J., Remy, N., and Scott, S. L. (2015). Inferring causal impact using Bayesian structural time-series models. *Annals of Applied Statistics*, 9(1), 247–274. https://doi.org/10.1214/14-AOAS788.

[7]. Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., et al. (2021). Variational quantum algorithms. *Nature Reviews Physics*, 3(9), 625–644. https://doi.org/10.1038/s42254-021-00348-9.

[8]. Chen, S., Cotler, J., Huang, H. Y., and Li, J. (2021). Exponential separations between learning with and without quantum memory. *arXiv preprint arXiv:2111.05881*. https://arxiv.org/abs/2111.05881.

[9]. Chowell, G., Viboud, C., Simonsen, L., and Moghadas, S. M. (2016). Characterizing the reproduction number of epidemics with early subexponential growth dynamics. *Journal of The Royal Society Interface*, 13(123):20160659.

[10]. Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance*, 1(2), 223–236.

[11]. Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd edition. Springer.

[12]. Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747), 209–212.

[13]. Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.

[14]. Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression:Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55–67.

[15]. Hoffman, M. D. and Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(1),1593–1623.

[16]. Huang, H. Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., and McClean, J. R. (2021). Power of data in quantum machine learning. *Nature Communications*, 12(1), 2631.

[17]. Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with Exponential Smoothing*. Springer.

[18]. Kadowaki, T. and Nishimori, H. (1998). Quantum annealing in the transverse Ising model. *Physical Review E*, 58(5):5355. https://doi.org/10.1103/PhysRevE.58.5355.

[19]. Kübler, J. M., Buchholz, S., and Schölkopf, B. (2021). The inductive bias of quantum kernels. *Advances in Neural Information Processing Systems*, 34, 12661–12673.

[20]. Liu, Y., Arunachalam, S., and Temme, K. (2021). A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9), 1013–1017.

[21]. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E., and Getz, W. M. (2005). Superspreading and the effect of individual variation on disease emergence. *Nature*, 438(7066), 355–359.

[22]. Mandelbrot, B. (1963). The variation of certain speculative prices. *The Journal of Business*, 36(4), 394–419.

[23]. McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., and Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1), 4812.

[24]. McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*, Revised edition Princeton University Press.

[25]. Schuld, M. and Killoran, N. (2019). Quantum machine learning in feature Hilbert spaces. *Physical Review Letters*, 122(4), 040504.

[26]. Schuld, M. and Petruccione, F. (2018). *Supervised Learning with Quantum Computers*. Springer. https://doi.org/10.1007/978-3-319-96424-9.

[27]. Schuld, M., Bocharov, A., Svore, K. M., and Wiebe, N. (2020). Circuit-centric quantum classifiers. *Physical Review A*, 101(3), 032308.

[28]. Scott, S. L. and Varian, H. R. (2014). Predicting the present with Bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisation*, 5(1-2), 4–23.

[29]. Somma, R., Boixo, S., Barnum, H., and Knill, E. (2008). Quantum simulations of classical annealing processes. *Physical Review Letters*, 101(13), 130504. https://doi.org/10.1103/PhysRevLett.101.130504.

[30]. Taylor, J. W. (2010). Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204(1), 139–152.

[31]. Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.

[32]. Weron, R. (2014). Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4), 1030–1081.

[33]. Zoufal, C., Lucchi, A., and Woerner, S. (2021). Quantum generative adversarial networks for learning and loading random distributions. *npj Quantum Information*, 7(1), 103. https://doi.org/10.1038/s41534-021-00448-1.