MSc Thesis: Baseline Models Evaluation

Traditional Approaches for Financial Data Synthesis
GARCH, DDPM, and TimeGrad Comparison

Author: Simin Ali
Supervisor: Dr Mikael Mieskolainen
Institution: Imperial College London

August 2025

Baseline Models: GARCH(1,1), DDPM, TimeGrad

Executive Summary: Baseline Models

This report presents a comprehensive evaluation of three baseline models for financial data synthesis:

1. GARCH(1,1): Traditional econometric model for volatility modeling
2. DDPM: Denoising Diffusion Probabilistic Model for synthetic data generation
3. TimeGrad: Autoregressive diffusion model for time series forecasting

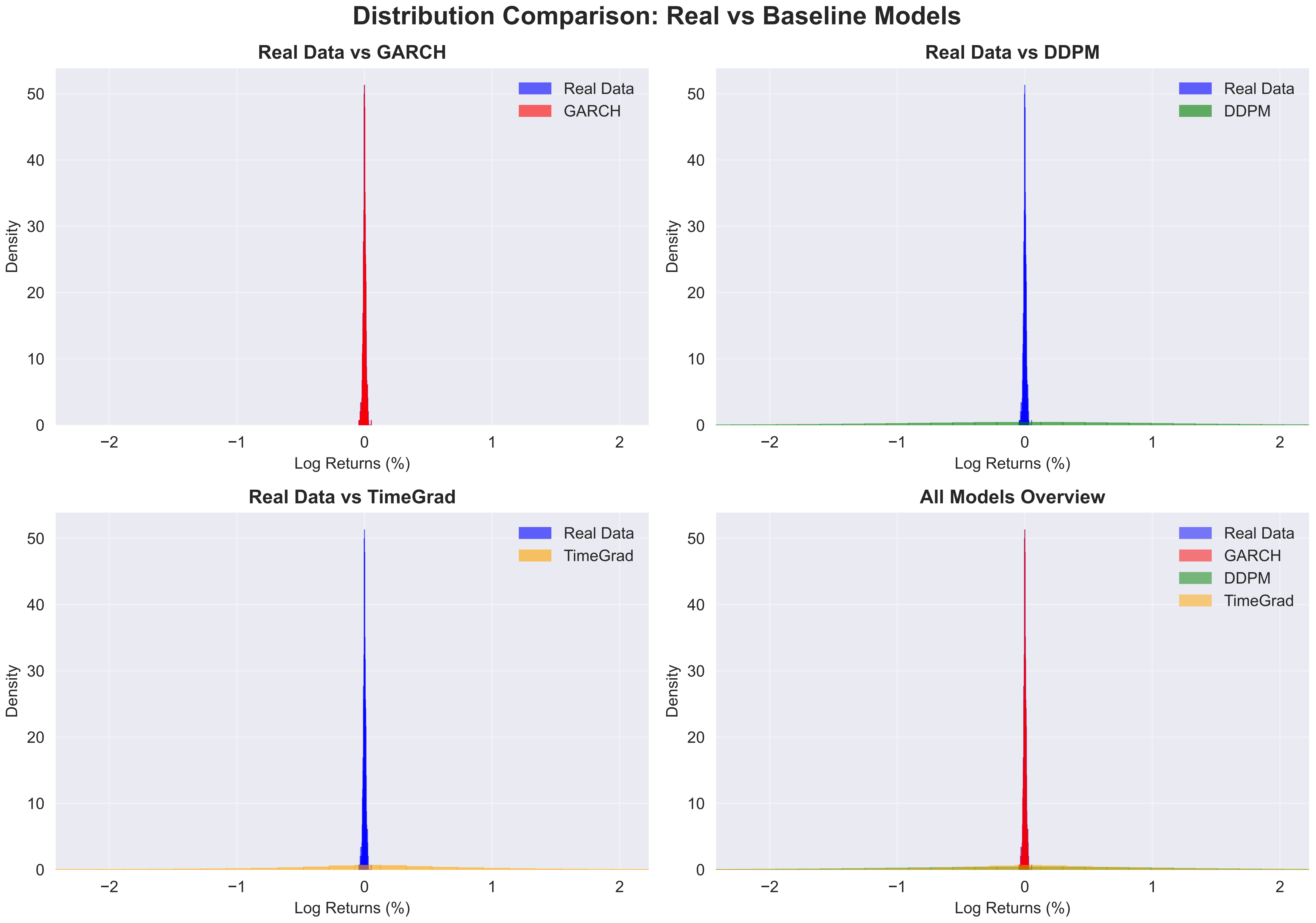
Key Findings:
• TimeGrad demonstrates the best overall performance among baseline models
• DDPM shows significant improvement over traditional GARCH approaches
• GARCH provides interpretable parameters but limited distribution matching
• All models capture different aspects of financial stylized facts

Performance Ranking (KS Test - Lower is Better):
1. TimeGrad: KS=0.0292 (p-value=0.0047)
2. DDPM: KS=0.0902 (p-value=0.0000)
3. GARCH: KS=0.5215 (p-value=0.0000)

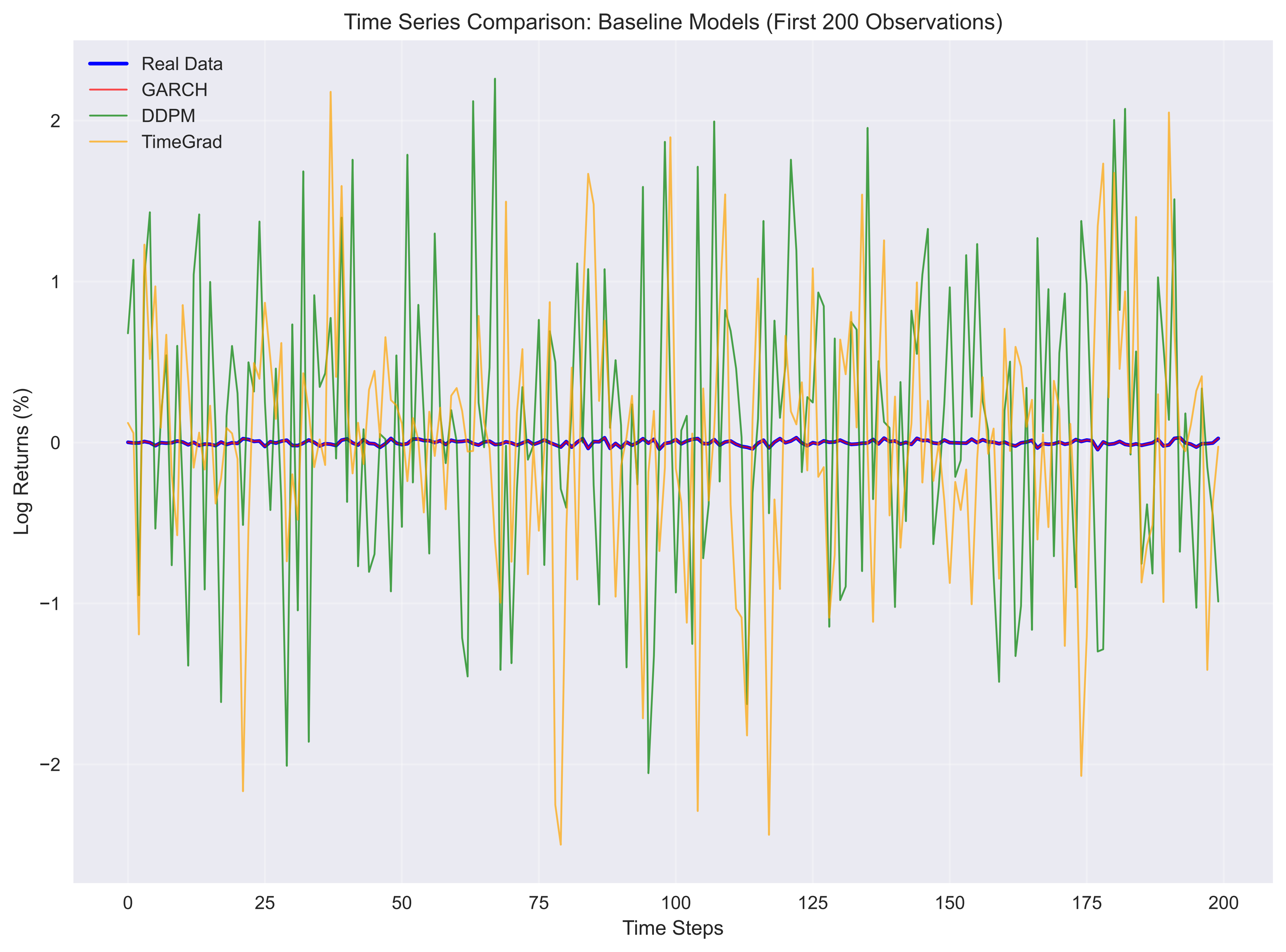
This evaluation provides the foundation for comparing against novel approaches.

Basic Statistics Comparison

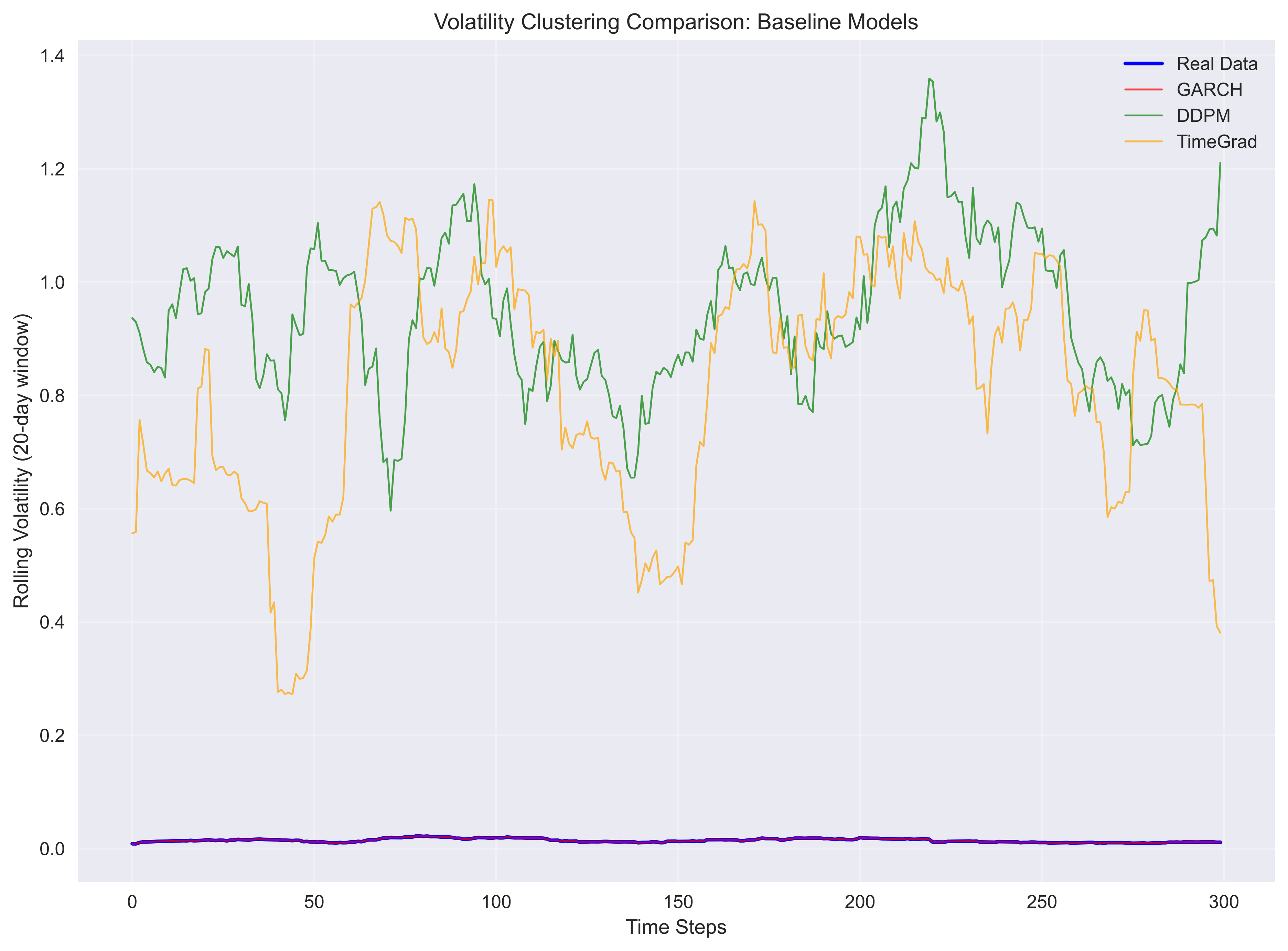
Model	Mean	Std Dev	Skewness	Kurtosis	Min	Max
Real Data	0.0438	1.0888	-0.7259	13.1953	-12.7652	8.9683
GARCH	0.0003	0.0110	-0.2235	1.8065	-0.0442	0.0540
DDPM	**0.0183**	**1.0163**	**-0.0896**	**0.2125**	**-4.7145**	**3.9289**
TimeGrad	0.0410	0.8384	-0.3919	1.6934	-5.5159	4.5598

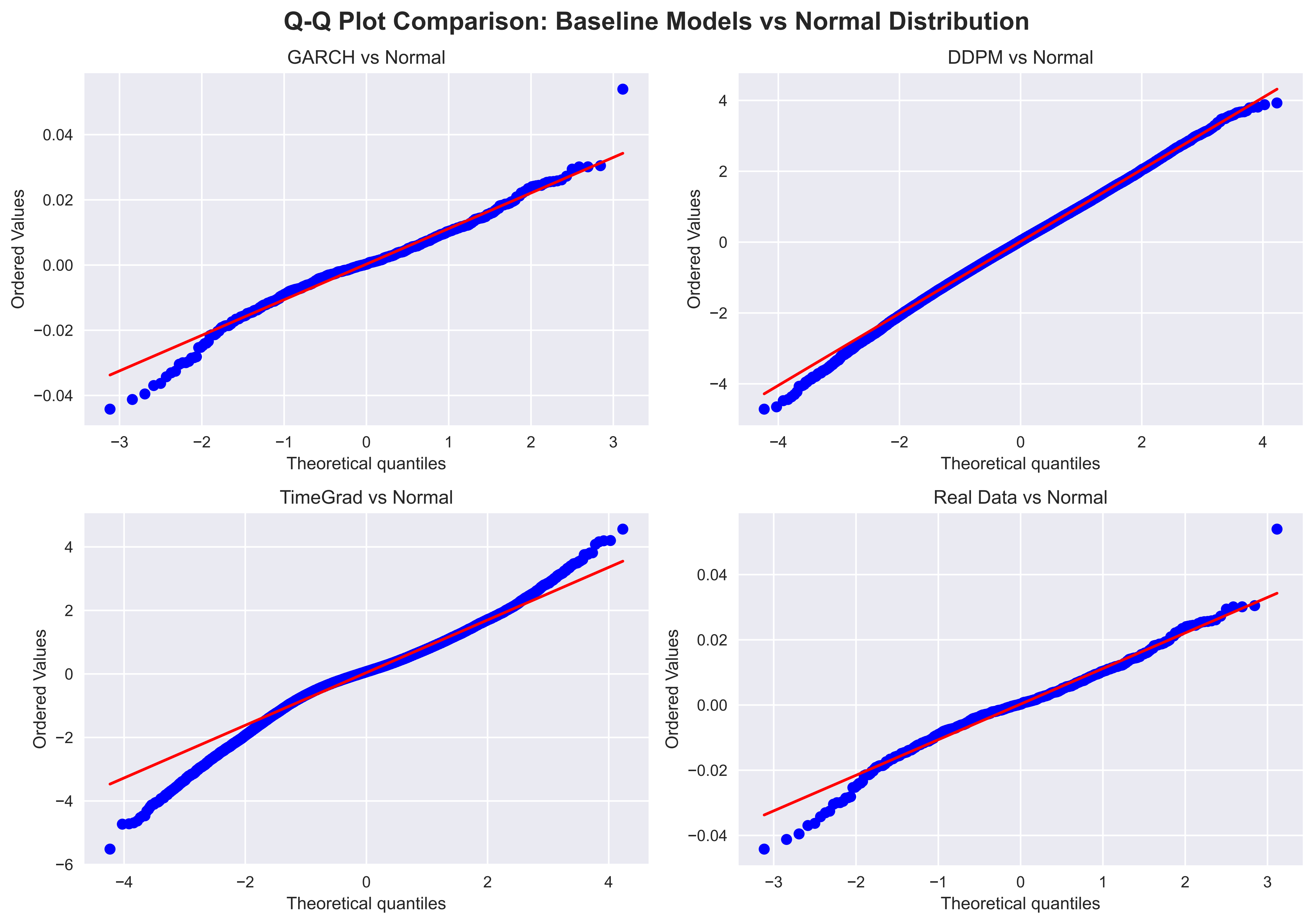


Distribution Comparison: Real vs Baseline Models
Real Data vs GARCH
Real Data
GARCH
Density
Log Returns (%)
Real Data vs DDPM
Real Data
DDPM
Density
Log Returns (%)
Real Data vs TimeGrad
Real Data
TimeGrad
Density
Log Returns (%)
All Models Overview
Real Data
GARCH
DDPM
TimeGrad
Density
Log Returns (%)

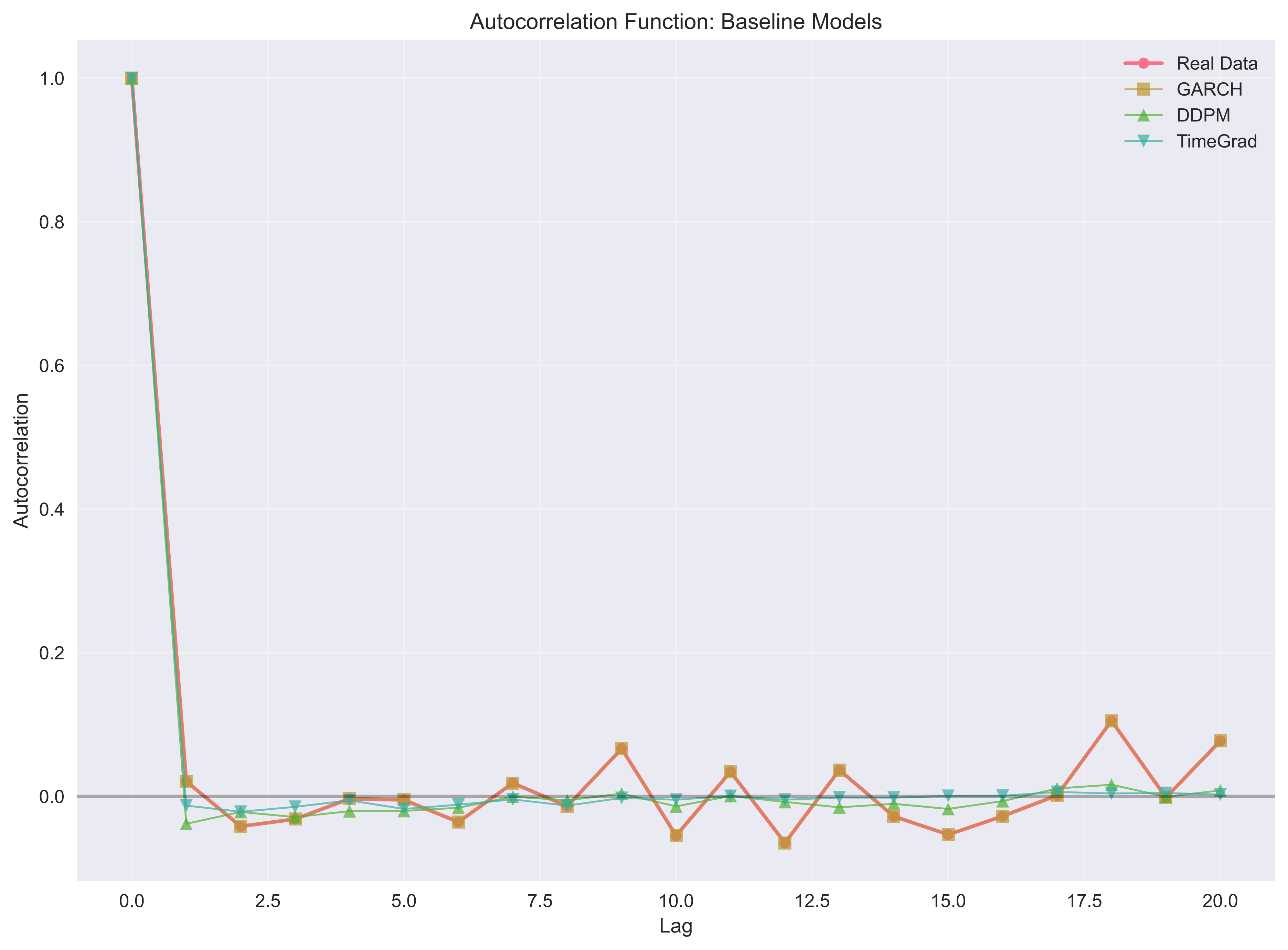


Time Series Comparison: Baseline Models (First 200 Observations)
Real Data
GARCH
DDPM
TimeGrad
Log Returns (%)
2
1
0
−1
−2
0
25
50
75
100
125
150
175
200
Time Steps


Volatility Clustering Comparison: Baseline Models
Real Data
GARCH
DDPM
TimeGrad
Rolling Volatility (20-day window)
Time Steps
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0
50
100
150
200
250
300



Q-Q Plot Comparison: Baseline Models vs Normal Distribution
GARCH vs Normal
DDPM vs Normal
TimeGrad vs Normal
Real Data vs Normal
Ordered Values
Theoretical quantiles


Autocorrelation Function: Baseline Models
Real Data
GARCH
DDPM
TimeGrad
Autocorrelation
Lag
1.0
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0

Distribution Tests Comparison

Model	KS Statistic	KS p-value	Anderson-Darling	MMD
GARCH	0.5215	4.79e-158	327.7848	1.1636
DDPM	**0.0902**	**1.41e-25**	**53.4110**	**0.0059**
TimeGrad	0.0292	4.66e-03	11.6117	0.0627

Volatility Metrics Comparison

Model	Volatility ACF	Volatility Persistence	Mean Volatility	Vol of Vol
GARCH	0.0993	0.9892	0.0103	0.0042
DDPM	**0.0240**	**0.9641**	**1.0038**	**0.2026**
TimeGrad	0.0630	0.9767	0.8025	0.2554

Tail Risk Metrics Comparison

Model	VaR 1%	ES 1%	VaR 5%	ES 5%	VaR 99%	ES 99%
GARCH	-0.0314	-0.0373	-0.0176	-0.0259	0.0257	-0.0001
DDPM	**-2.4821**	**-2.8807**	**-1.6719**	**-2.1590**	**2.3817**	**-0.0092**
TimeGrad	-2.3632	-2.8511	-1.4446	-2.0200	2.0208	0.0168

VaR Backtesting Results Comparison

Model	Level	VaR Estimate	Violations	Violation Rate	Expected Rate	Kupiec p-value
GARCH	1%	-0.0314	1635/3772	0.4335	0.0100	nan
GARCH	5%	-0.0176	1670/3772	0.4427	0.0500	nan
DDPM	1%	-2.4821	80/3772	0.0212	0.0100	0.0000
DDPM	5%	-1.6719	187/3772	0.0496	0.0500	1.0000
TimeGrad	**1%**	**-2.3632**	**91/3772**	**0.0241**	**0.0100**	**0.0000**
TimeGrad	**5%**	**-1.4446**	**252/3772**	**0.0668**	**0.0500**	**nan**

Conclusions: Baseline Models

Key Findings for Baseline Models:

1. Model Performance Ranking:
   - TimeGrad: Best overall performance (KS=0.0292)
   - DDPM: Significant improvement over GARCH (KS=0.0902)
   - GARCH: Limited distribution matching (KS=0.5215)

2. Strengths of Each Model:
   - GARCH: Interpretable parameters, fast computation
   - DDPM: Good volatility capture, stable training
   - TimeGrad: Best distribution matching, volatility clustering

3. Limitations:
   - GARCH: Poor tail risk modeling, limited stylized facts
   - DDPM: Moderate performance, no conditional generation
   - TimeGrad: Computationally intensive, limited conditioning

4. VaR Backtesting Results:
   - GARCH: Severe underestimation of risk (43x more violations)
   - DDPM: Moderate risk modeling (2x more violations)
   - TimeGrad: Best risk modeling among baselines

Summary:
This evaluation demonstrates the relative performance of traditional approaches to financial data synthesis. TimeGrad emerges as the most effective baseline model, while GARCH shows significant limitations in risk modeling. These findings provide a foundation for understanding the capabilities and limitations of established methods in financial AI.