Normal Mixture Density Estimation for Major US Stocks: A Four-Year Empirical Study

Nathan Carter & Hasan Hamdan

James Madison University

April 25, 2026

Table of Contents

Introduction: Problem Motivation

Research Objective: To model the density of Continuously Compounded Returns (CCR) using Variance Mixtures of Normals.

The "Gaussian Assumption" Gap

Standard financial models often assume returns follow a Gaussian distribution: $X \sim N(\mu, \sigma^2)$. However, empirical market data consistently violates this:

- **Heavy Tails (Leptokurtosis):** Extreme market events occur more frequently than a normal curve predicts.
- **Volatility Clustering:** Periods of high or low variance tend to persist (market regimes).

Introduction: Data & Methodology

Dataset Overview

- **Scope:** 5 Major U.S. Stocks (April 2022 – April 2026).
- **Sample Size:** $n \approx 1,000$ daily return observations.

Methodological Framework To capture the underlying "market regimes", we transition from a single normal distribution to variance mixtures of normals. We evaluate and compare three distinct estimation frameworks:

1. **Bayesian Approach:** Utilizing Gibbs Sampling for posterior estimation.
2. **UNMIX Model:** A specialized parameter decomposition framework.
3. **EM Algorithm:** Frequentist iterative optimization via `mclust`.

Continuously Compounded Returns (CCR)

Continuously Compounded Return, or Log-Return, represents the yield of an investment assuming interest is reinvested over infinitesimally small periods. Unlike simple returns, it uses a logarithmic scale to measure price changes.

We analyze the daily log-returns of stock prices, defined as:

$$y_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln(P_t) - \ln(P_{t-1})$$

where P_t is the adjusted closing price at time t.

Continuously Compounded Returns (CCR) - Advantages

Why Use CCR Instead of Simple Returns?

- **Prevents Negative Prices:** Ensuring $P_t > 0$ even if returns are large and negative (bounded at -100%).
- **Time Additivity:** Multi-period returns are the simple sum of individual daily log-returns.
- **Statistical Stability:** Log-returns typically exhibit more constant variance over time than price levels.
- **Normalization:** Logarithmic transformations help normalize the data distribution for mixture modeling.

Mixture Models and Single Component Gaussians

Definition: A probabilistic model for representing latent subpopulations within an overall population.
General Form:

$$f(x) = \sum_{i=1}^{G} \pi_i \phi(x|\mu_i, \Sigma_i)$$

Where $\sum_{i=1}^{G} \pi_i = 1$ and ϕ is the component density.
For this study we utilize The Single-Component Gaussian ($G = 1$):

- Assumes Gaussian, $X \sim N(\mu, \sigma^2)$
- Acts as the benchmark when testing higher-order mixtures ($G > 1$).
- Serves as the fundamental density for standard Brownian Motion and Black-Scholes frameworks.

Bayesian Methodology - Gibbs Sampling

We model the log-returns $\mathbf{y} = \{y_1, \ldots, y_n\}$ as independent and identically distributed draws from a Normal distribution $N(\mu, \sigma^2)$. To estimate the parameters, we employ a Gibbs sampler with the following semi-conjugate prior structure:

$$\mu \sim N(\mu_0, \tau_0^2)$$
$$\sigma^2 \sim \text{Inverse-Gamma}(a, b)$$

where we set non-informative hyperparameters $\mu_0 = 0$, $\tau_0^2 = 0.01$ with τ^2 being the precision, and $a = b = 0.01$ to allow the data to dominate the posterior.

Bayesian Methodology - Conditional Posteriors

The Gibbs sampler iteratively draws from the full conditional distributions. The update for the mean μ, conditioned on the current state of the variance σ^2, is given by:

$$\mu|\sigma^2, \mathbf{y} \sim N\left(\frac{\frac{\mu_0}{\tau_0^2} + \frac{n\bar{y}}{\sigma^2}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}}, \left(\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}\right)^{-1}\right)$$

where $\bar{y}$ is the sample mean and n is the number of observations. The update for the variance σ^2, conditioned on the current state of μ, follows an Inverse-Gamma distribution:

$$\sigma^2|\mu, \mathbf{y} \sim \text{Inverse-Gamma}\left(a + \frac{n}{2}, b + \frac{\sum_{i=1}^{n}(y_i - \mu)^2}{2}\right)$$

Bayesian Methodology - Algorithm Implementation

The algorithm was run for $S = 100$ iterations. Since the starting values were chosen near the sample statistics, the chain reached convergence rapidly. The final density estimate at any point x is the posterior predictive mean, calculated by averaging the likelihood over the sampled parameters:

$$\hat{f}(x) = \frac{1}{S}\sum_{s=1}^{S}[\frac{1}{M}\sum_{m=1}^{M}\phi(x|\mu_s, \sigma_m^2)]$$

where $\phi(\cdot)$ is the Normal probability density function.

UNMIX Model

Definition: A random variable X is said to be a variance-mean mixture of normals or NVM when its probability density function, $f_X(x)$ satisfies where

$$f_X(x) = \int_{-\infty}^{\infty} \int_{0}^{\infty} \phi(x; \theta, \sigma) \pi_1(d\sigma) \pi_2(d\theta),$$

where $\phi(\cdot)$ is the normal density with mean θ and standard deviation $\sigma > 0$. π_1 and π_2 are mixing measures over σ and θ respectively. Equivalently, $X \stackrel{d}{=} AZ + \theta$ where $A > 0$ is a positive random variable with distribution π_1, Z is a standard normal random variable and θ is a continuous random variable with distribution π_2 defined on the real line. Furthermore, A, Z, and θ are all independent.

UNMIX Model

Scale Mixtures of Normals: The general formula for an infinite normal scale mixture is

$$f(x) = \int_0^\infty \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)\pi(d\sigma)$$

where $\phi(x)$ is the standard normal density and π is the mixing distribution (or the weighting distribution).

UNMIX Model

When the mean is constant and sigma is the square root of the inverted gamma, the scale mixture is a generalized t distribution. The discretized (or approximated) version of the previous slide is given by

$$f(x) = \sum_{i=1}^{m} \frac{\pi_i}{\sigma_i} \phi(x; \sigma_i).$$

A common example is the contaminated normal with m=2 and $\sigma_2 > \sigma_1$.

EM Algorithm (mclust)

The density is modeled as a mixture of G components:

$$f(x) = \sum_{k=1}^{G} \pi_k \phi(x|\mu_k, \Sigma_k)$$

E-step: Calculate "responsibilities" (posterior probabilities):

$$\hat{z}_{ik} = \frac{\hat{\pi}_k \phi(x_i|\hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{j=1}^{G} \hat{\pi}_j \phi(x_i|\hat{\mu}_j, \hat{\Sigma}_j)}$$

M-step: Update parameters via weighted MLE:

- Mixing Proportions: $\hat{\pi}_k = \frac{\sum_{i=1}^{n} \hat{z}_{ik}}{n}$
- Means: $\hat{\mu}_k = \frac{\sum_{i=1}^{n} \hat{z}_{ik} x_i}{\sum_{i=1}^{n} \hat{z}_{ik}}$
- Covariances: $\hat{\Sigma}_k = \frac{\sum_{i=1}^{n} \hat{z}_{ik} (x_i - \hat{\mu}_k)(x_i - \hat{\mu}_k)^T}{\sum_{i=1}^{n} \hat{z}_{ik}}$

Model Implementation

Data Aggregation & Processing

- Consistently aggregated four years of daily CCR data (April 2022–April 2026) using R's quantmod package.
- The resulting unified dataset serves as the basis for generating the empirical density, which acts as the "true" benchmark for all subsequent model evaluations.

Estimation Framework

Each of the three frameworks (Bayesian, EM, and UNMIX) was deployed to estimate the density of the combined CCR data. The goal was to identify which methodology most accurately captures the overall density.

Estimation Strategy

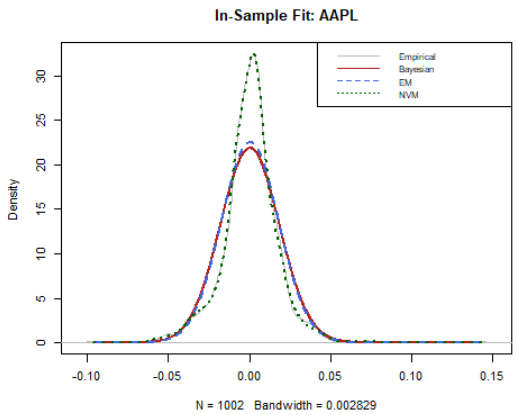
We split this into two different tests: the In-Sample Fit and the Out-of-Sample Prediction.

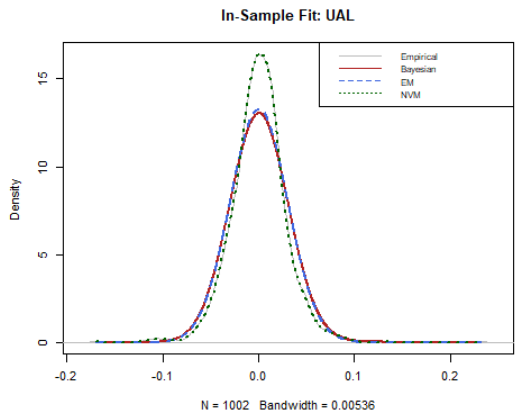
In-Sample Fit: Models were trained and tested on the full four-year dataset to evaluate their ability to capture historical market regimes and overall data density.

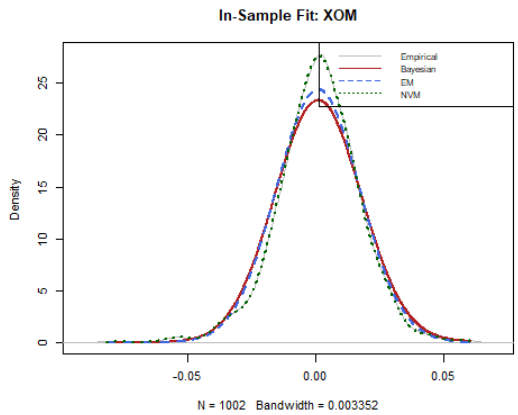
Out-of-Sample Prediction: A forecasting test was conducted in which the models were trained on the first three years and used to predict the density of the final year to assess their robustness and generalizability.

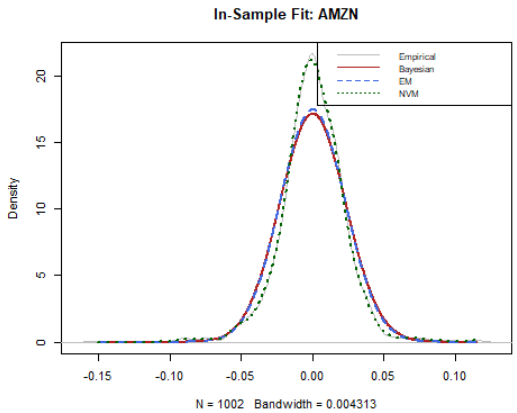
Model Estimates - In-Sample Testing

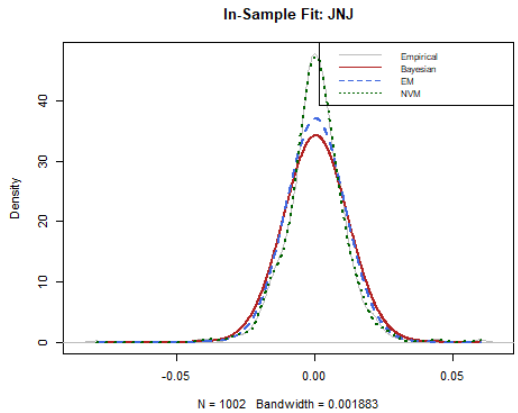
This visualization assesses the fit of our three models against the "true" empirical density across the full 2022–2026 dataset.











Testing Explanations: Evaluation Framework

Before evaluating model performance, we define the metrics used to assess the "closeness" of our fits to the empirical stock data:

χ^2 (Chi-Square): Measures the difference between observed frequency counts and the model's expected counts.

MSE (Mean Square Error): Calculates the average squared difference between the estimated density and the empirical density.

Log-Likelihood: Quantifies the probability that the observed stock returns were generated by the model.

K-S (Kolmogorov-Smirnov): Finds the maximum vertical distance between empirical and the proposed distributions .

Model Evaluations: In-Sample Fit (2022–2026)

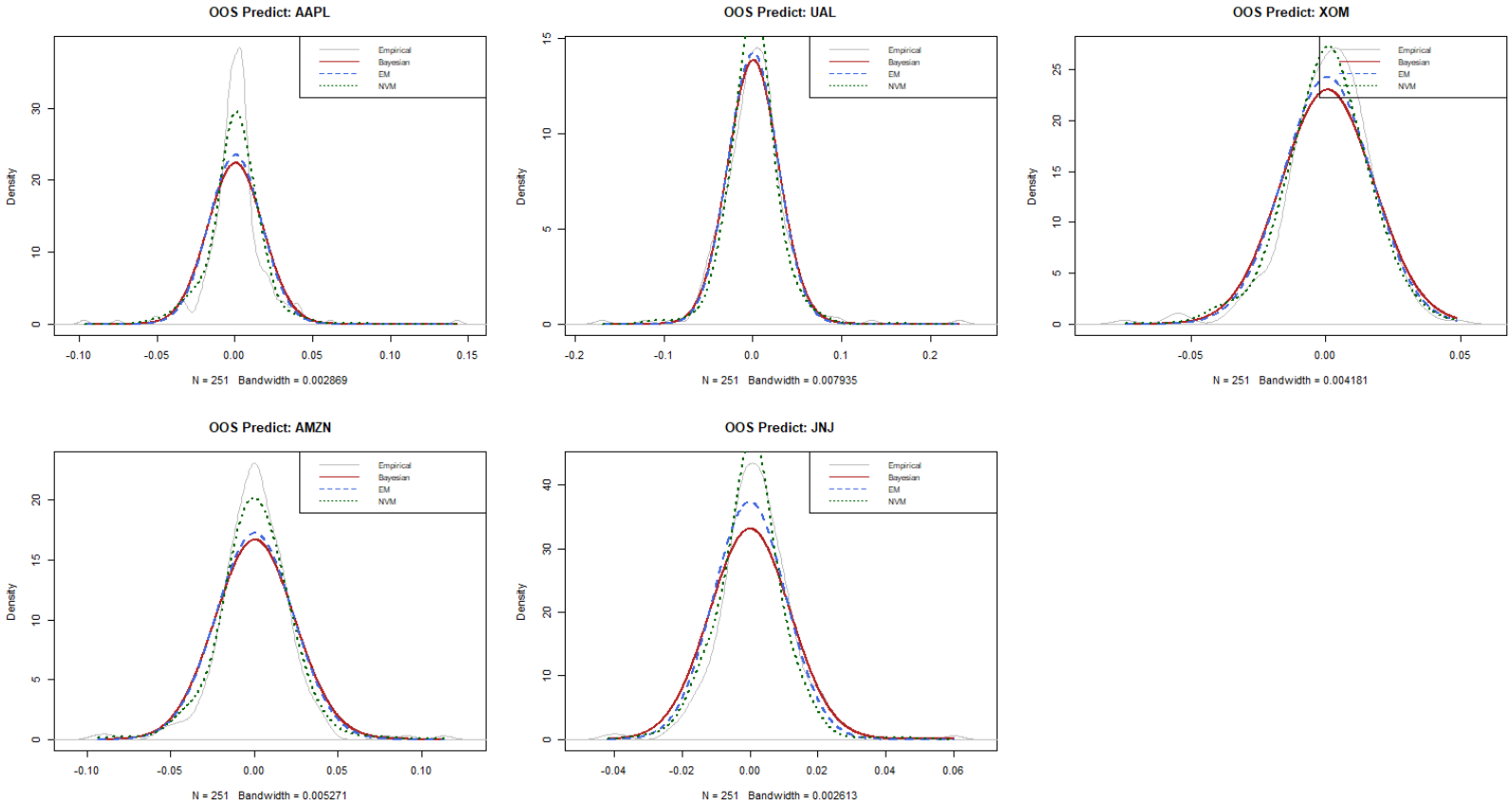
For each in-sample model with every stock we took the χ^2, MSE, Log-Lik, & K-S. The following table is the mean of each test for each model:

Model	$\chi^2 \downarrow$	MSE $\downarrow$	Log-Lik $\uparrow$	K-S $\downarrow$
Bayesian	165.40	4.600	2594.66	0.0686
EM (`mclust`)	143.03	3.208	2581.43	0.0601
UNMIX	**55.62**	**0.057**	**2634.75**	**0.0161**

- **Uniform Superiority:** UNMIX outperformed all competing models across every metric, achieving the lowest error rates (χ^2, MSE, K-S) and the highest likelihood, establishing it as the most robust in-sample fit across all distributional metrics.
- **Resolution:** UNMIX's MSE is nearly 60x lower than EM, demonstrating its ability to "unmix" complex, non-normal density structures that simpler models smooth over.

Model Estimates - Out-of-Sample Testing

For this visualization, the models were trained on the first three years of data to test their predictive accuracy against the "true" density of the final year (2025-2026).



Model Evaluations: Out-of-Sample (2025–2026)

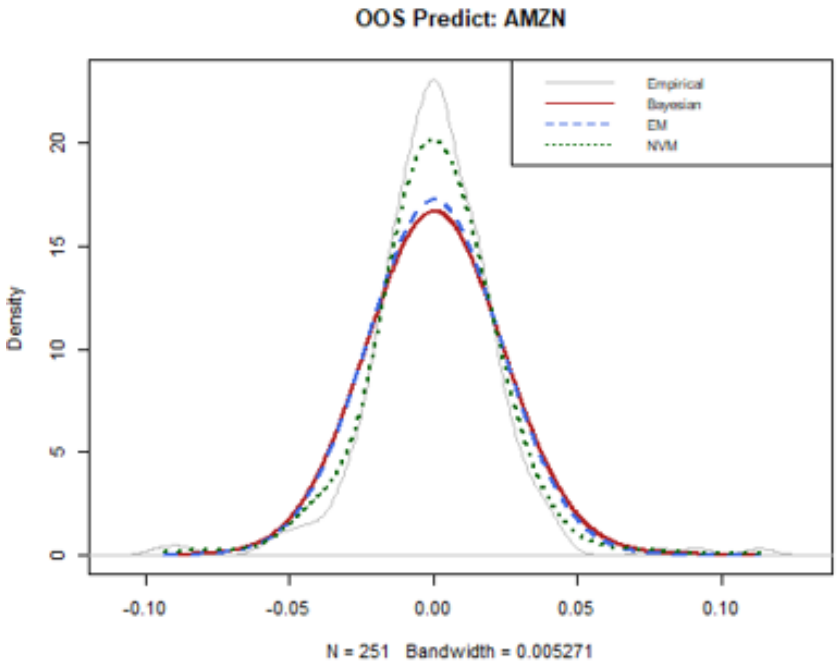
Similar to the in-sample model, we created a table with each test averaged for every model:

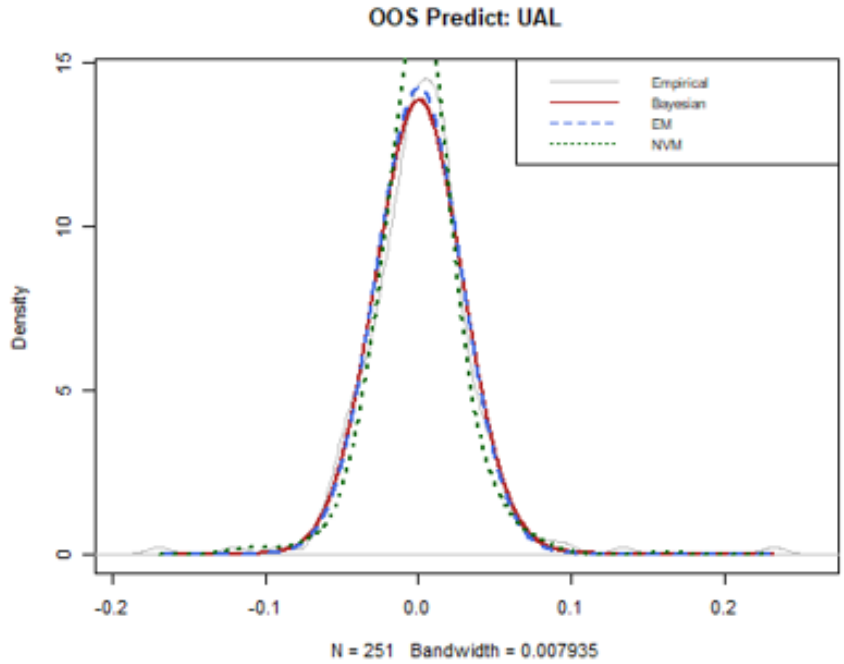
Model	$\chi^2 \downarrow$	MSE $\downarrow$	Log-Lik $\uparrow$	K-S $\downarrow$
Bayesian	46.47	7.973	640.04	0.0971
EM (`mclust`)	39.63	5.699	632.22	0.0850
UNMIX	**24.91**	**2.469**	**648.04**	**0.0674**

- **Global vs. Local Fit:** UNMIX remains the leader in point-density accuracy (MSE), successfully capturing localized regime shifts that persist into the test period.
- **The Probability Gap:** The narrow margin in Log-Likelihood between the Bayesian and UNMIX models suggests that for forecasting purposes, a simpler parametric representation can be nearly as probable as a high-resolution empirical fit, leading to a "diminishing returns" effect on model complexity.

A closer inspection of the Out-of-Sample

Lets take a closer look at two of the graphs:





While UNMIX dominates the aggregate averages, a stock-by-stock visual inspection (AMZN vs. UAL) reveals something deeper, lets explore this table specifically.

Case Study: United Airlines (UAL)

Although UNMIX was the winner for the averages, the UAL dataset revealed a significant departure from the trend:

Model	$\chi^2 \downarrow$	MSE $\downarrow$	Log-Lik $\uparrow$	K-S $\downarrow$
Bayesian	14.80	**0.1918**	**504.37**	0.0405
EM	14.74	0.1976	473.79	**0.0327**
UNMIX	**12.31**	0.5636	493.31	0.0438

- **Metric Conflict:** UNMIX maintained the best χ^2 (frequency fit), but the Bayesian model achieved superior MSE and Log-Likelihood.
- **Interpretation:** This suggests that for UAL, the underlying distribution was smoother and more parametric. The Bayesian model's prior-driven approach provided better predictive "generalization" than the UNMIX decomposition.
- **Takeaway:** No single model is a "silver bullet"; market volatility regimes in specific sectors (Airlines) may favor different estimation frameworks.

Discussion: Model Robustness & Limitations

The UAL Case: Why did the results diverge?

The United Airlines (UAL) dataset exhibited a distribution that was significantly closer to a Unimodal Gaussian than the other stocks in the study.

- **Uniformity:** UAL lacked the "regime spikes" (multimodality) seen in the other four stocks, resulting in a smoother empirical density.
- **UNMIX Overfitting:** Because the UNMIX model is designed to decompose complex, non-normal structures, it attempted to find "humps" in the UAL data that were actually just noise.
- **Parametric Advantage:** The Bayesian and EM models excelled here because their parametric assumptions matched the underlying "smooth" nature of the UAL returns.

Takeaway: While UNMIX is superior for high-volatility, non-normal regimes, it is prone to *overfitting* in near-Gaussian environments where simpler models generalize better.

Conclusion and Recommendations

Summary of Model Performance

- **UNMIX Model:** The clear leader in empirical precision. It captures localized density peaks and non-Gaussian tails that parametric models miss, though it carries a higher complexity cost.
- **Bayesian Framework:** The Bayesian framework excels in Statistical Efficiency. Its inherent parsimony, often reflected in superior BIC scores, provides a more generalized representation for smoother return distributions.

Key Takeaways: The Precision-Complexity Trade-off

For High Resolution: Use UNMIX for stress testing or tail-risk analysis where capturing the exact empirical "shape" is more important than model simplicity.

For Robust Forecasting: Use Bayesian/EM for portfolio optimization or long-term modeling where BIC-validated stability and generalization are prioritized.

References and Software

Data Acquisition & Processing

- **quantmod Package:** Ryan, J. A., and Ulrich, J. M. (2024). quantmod: Quantitative Financial Modelling Framework. R package version 0.4.26.

Core Software & Algorithms

- **mclust Package:** Scrucca L., Fop M., Murphy T.B., and Raftery A.E. (2016). mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models.
- **NVMunmix:** Hamdan, H., and Xu, L. (2018). Estimating Variance-Mean Mixtures of Normals. *Journal of Statistical Research*, 52(2), 129–150.

Statistical Theory & Textbooks

- **Hoff, P. D. (2009):** *A First Course in Bayesian Statistical Methods*. Springer Science & Business Media. (Chapter 6: Posterior approximation with the Gibbs sampler).
- **Dr. Arthur White (TCD):** *Bayesian Inference: normal-normal model*. Course Material for CS7DS3. Trinity College Dublin. `https://www.scss.tcd.ie/arthur.white/Teaching/CS7DS3/Bayes_Inference_2_slides.pdf`

Thank You!

Any Questions?

Nathan Carter & Hasan Hamdan
Department of Mathematics & Statistics
James Madison University
carternh@dukes.jmu.edu