Robust Estimation in Autoregressive Models with Skewed Innovations

Mohamed S. Mohamed[1], Adrienne Katner[2], Barry Keim[2], Evrim Oral[1*]

LSUHSC School of Public Health, [1]BSDS Program [2] EHCS Program

LSU Health NEW ORLEANS School of Public Health

Introduction

In time series analysis, autoregressive (AR) models in the form:

$$y_i = \sum_{j=1}^{q} \phi_j y_{i-j} + a_i \qquad (1)$$

Traditionally uses estimation methods such as Ordinary Least Squares (OLS) and Yule-Walker (YW), which rely on the assumption of normally distributed innovations (a_i). This theoretical assumption may not hold true in empirical reality. In diverse real-world applications, time series data can exhibit severe skewness, heavy tails, and structural outliers resulting in significantly biased and statistically inefficient estimates.

Several robust alternatives have been used in other statistical models such as M-estimators; however, their performance in time-series models with innovations following skewed distribution was not evaluated. Unlike other robust estimators, the Modified Maximum Likelihood (MML) method builds on the correct assumption of the innovation distribution and utilizes ordered statistics and Taylor series expansions to offers a robust, computationally tractable alternative. This project extends the application of MML estimators to AR models (with generalization to high order AR(q) models) under right-skewed distributed innovations.

Methods

Skewed Distributions We discuss the scenarios where the innovations a_i in equation (1) violates normality and follows skewed distributions: (1) Inverse Gamma (2) Log-Normal (Figure 1)

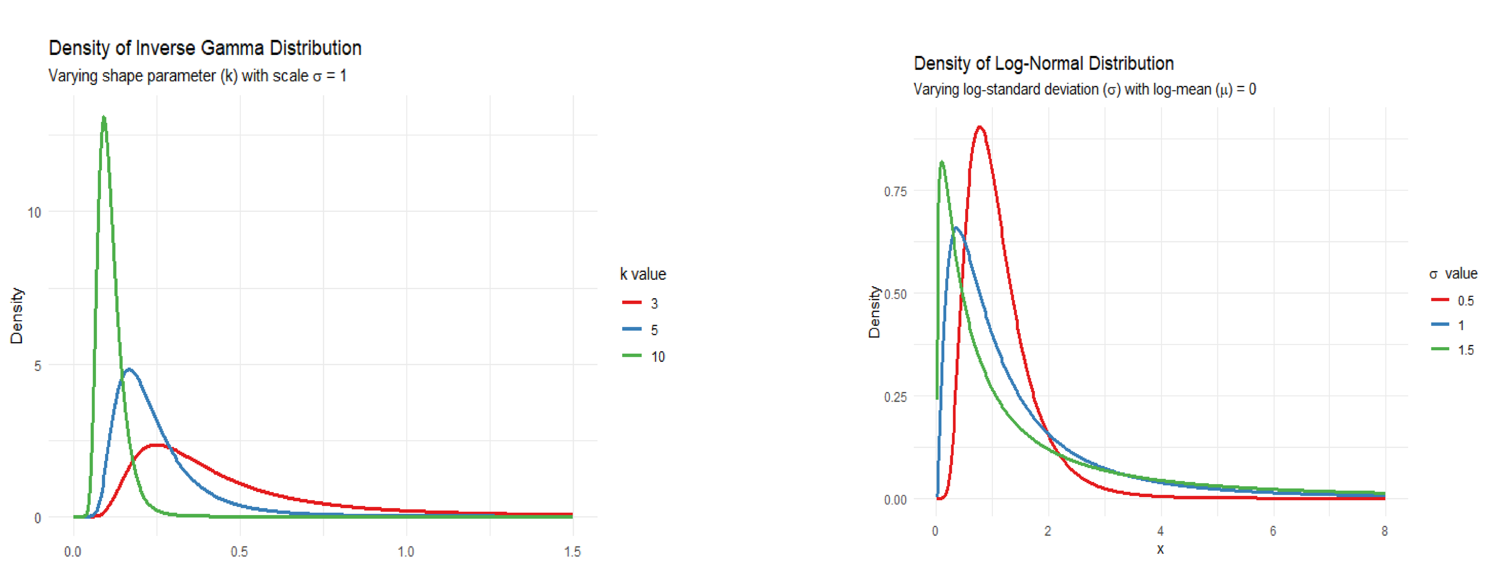

Figure 1. Density function of the Inverse gamma distribution (left) and Log-Normal distribution (right), under different shape and scale parameters.

MML Method: We generalized MML method to these distributions through the following steps: (1) Driving The Likelihood Functions, (2) Using Ordered Statistics, (3) Taylor series expansion (4) Deriving and Maximizing The "Modified" Likelihood Function. We compared **MML** method to traditional estimators: **OLS** and **YW**, as well as robust estimators: **GM**-estimators, **MM**-estimators and Least Absolute Deviation Estimators (**LAD**).

Simulation Settings: The settings in our study are summarized in Table 1.

Table 1. Simulations settings for Inverse Gamma and Log-Normal distributed innovations

Specifications	Inverse Gamma	Log-Normal
Sample size (n)	*30, 100, 1000*	*30, 100, 1000*
Autoregressive Parameters (ϕ)	**S1**: $\phi = 0.5$, **S2**: $\phi_1 = 0.6, \phi_2 = -0.2$, **S3**: $\phi_1 = 0.9, \phi_2 = -0.5, \phi_3 = 0.2$	**S1**: $\phi = -0.5$, **S2**: $\phi = 0$, **S3**: $\phi = 0.1$, **S4**: $\phi = 0.5$, **S5**: $\phi = 0.9$
Distribution specific parameters	Shape parameter $k = 3, 5, 10$	Scale Parameter $\sigma = 0.5, 1, 1.5$

Results

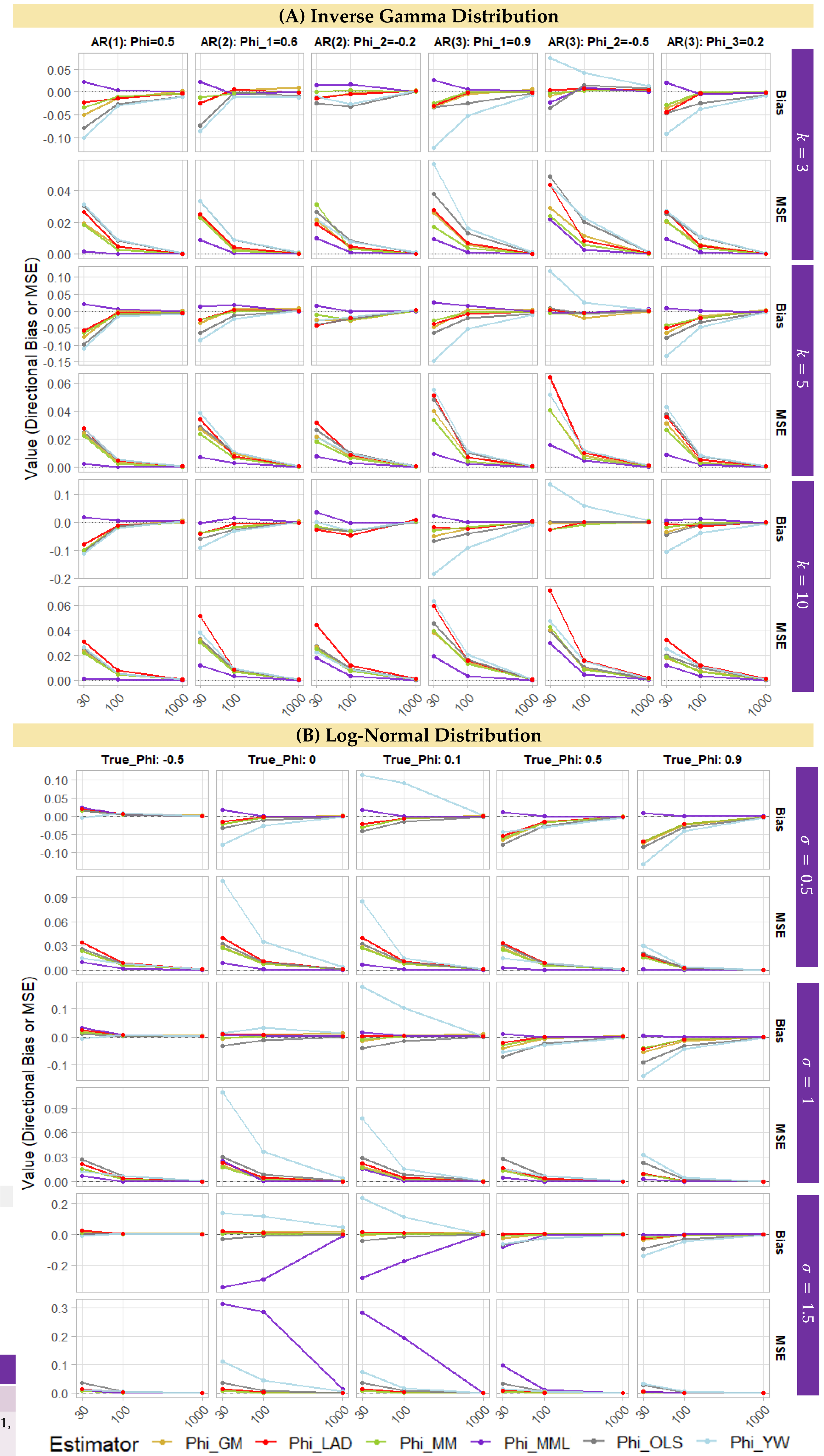

Figure 2. Bias and Mean Squared Error (MSE) results for (A) Inverse Gamma and (B) Log-Normal Distributed Innovations under different simulation settings.

Results (cont.)

- MML method shows superior performance in most of the settings compared to both traditional estimators and robust estimators.
- Traditional estimators (OLS and YW) are not reliable and should not be used in with skewed innovations, especially in low and moderate sample sizes. Increasing the sample size (≥ 1000) can be a remedy for their disadvantage.

Inverse Gamma Innovations: MML method didn't only outperformed other estimators but also was *the only unbiased and efficient* across all simulation settings. Traditional estimators were biased in low and moderate sample sizes. Robust estimators were unbiased; however, they were not efficient in small sample sizes.

Log-Normal Innovations: In most scenarios MML outperformed all other estimators. However, when the scale parameter is large ($\sigma > 1$), MM-estimator outperformed all other estimators, and MML was biased for small sample sizes. In 12 of 15 scenarios, MML method performance was better than traditional estimators.

Table 2. Best performing estimator in different simulation settings as measured by Bias and Mean Squared Error

(A) Inverse Gamma Distribution

Scenario	S1: $\phi = 0.5$			S2: $\phi_1 = 0.6, \phi_2 = -0.2$			S3: $\phi_1 = 0.9, \phi_2 = -0.5, \phi_3 = -0.5$		
n	30	100	1000	30	100	1000	30	100	1000
$k = 3$	MML	MML	MML	MML	MML	MML	MML	MML	MML
$k = 5$	MML	MML	MML	MML	MML	MML	MML	MML	MML
$k = 10$	MML	MML	MML	MML	MML	MML	MML	MML	MML

(B) Log-Normal Distribution

Scenario	S1: $\phi = -0.5$			S2: $\phi = 0$			S3: $\phi = 0.1$			S4: $\phi = 0.5$			S5: $\phi = 0.9$		
n	30	100	1000	30	100	1000	30	100	1000	30	100	1000	30	100	1000
$\sigma = 0.5$	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML	MML
$\sigma = 1$	MML	MML	MML	MM	MM	MM	MML	MML	MML	MML	MML	MML	MML	MML	MML
$\sigma = 1.5$	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM	MM

Conclusion/Future Work

- In AR(q) models with skewed innovations, MML is superior to conventional and robust estimators across all Inverse gamma innovation settings and most Log-normal innovation settings.
- MM-estimators performed better in Log-normal innovations with large scale parameter (more skewness).
- Conventional estimators were biased and shouldn't be used with small and moderate sample sizes.
- Our ultimate goal is to apply our estimators to estimate time series from environmental data

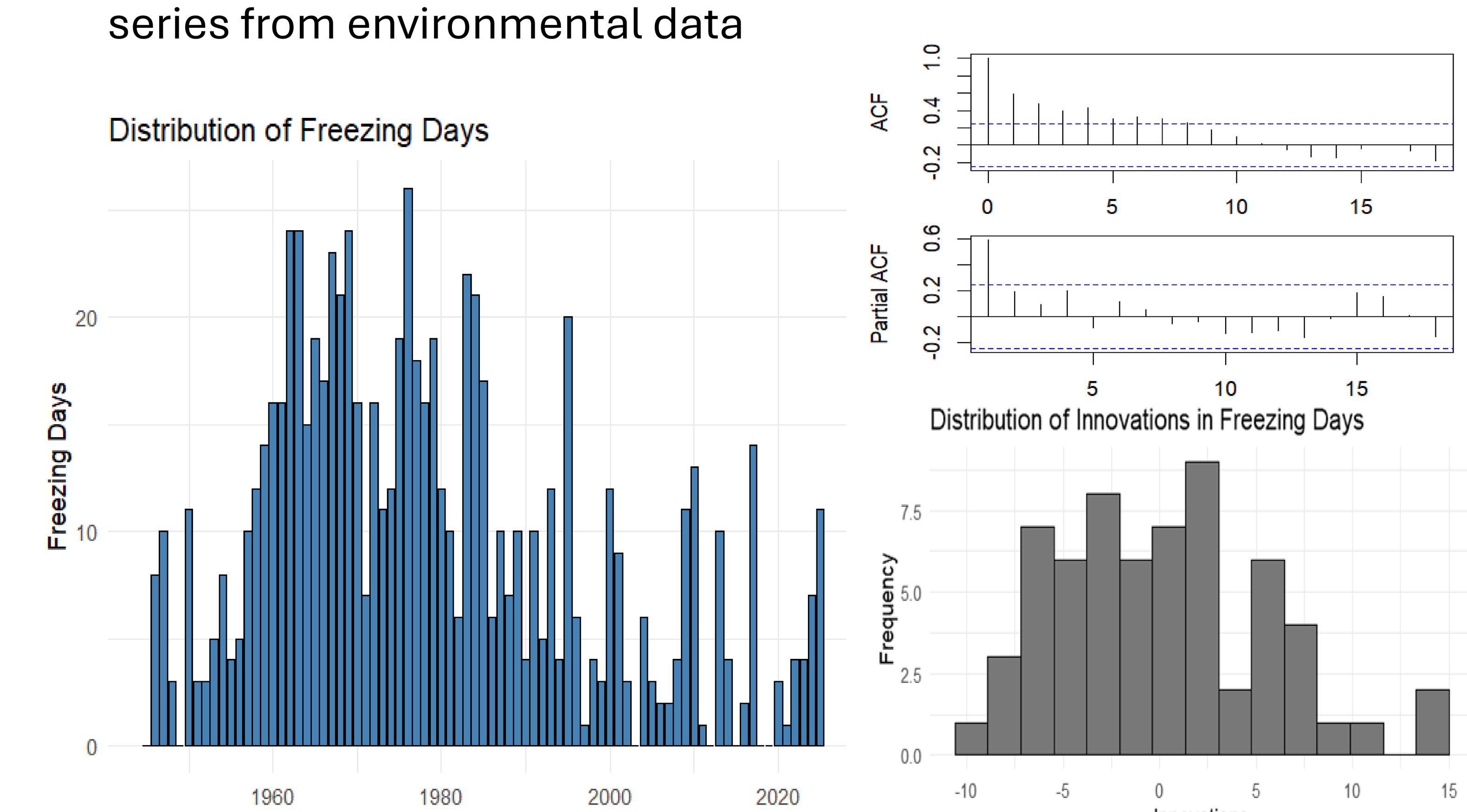

Figure 3. Distribution of Freezing Days in New Orleans, LA and AR(1) model Innovations