EFFICIENT ESTIMATION FOR TIME SERIES MODELS

A Thesis

Submitted to the Faculty of Graduate Studies and Research

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

In

Statistics

University of Regina

By

Sharandeep Singh Pandher

Regina, Saskatchewan

January, 2023

©Copyright 2023: S. S. Pandher
UNIVERSITY OF REGINA

FACULTY OF GRADUATE STUDIES AND RESEARCH

SUPERVISORY AND EXAMINING COMMITTEE

Sharandeep Singh Pandher, candidate for the degree of Doctor of Philosophy in Statistics, has presented a thesis titled, Efficient estimation for time series models, in an oral examination held on December 23, 2022. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: Dr Umberto Triacca, Università degli Studi dell'Aquila *

Supervisor: Dr Shakhawat Hossain, Adjunt, Department of Mathematics and Statistics *

Committee Member: Dr Andrei Volodin, Department of Mathematics and Statistics*

Committee Member: Dr DianLiang Deng, Department of Mathematics and Statistics*

Committee Member: Dr Arzu Sardarli, Associate, Department of Mathematics and Statistics *

Committee Member: Dr Yiyu Yao, Department of Computer Science *

Chair of Defense: Dr Chris Oriet, Faculty of Graduate Studies and Research *

*Attended via Zoom conferencing
Abstract

In this dissertation, we discuss the pretest, improved pretest and shrinkage estimators for some time series regression models, such as ARIMA, GARMA, and GARFIMA. We study the asymptotic biases and risks of these estimators and compare their relative performance with respect to the maximum likelihood estimator (MLE) analytically and numerically using Monte Carlo experiments and real data examples.

In Chapter 2, we consider the James-Stein shrinkage estimation procedure for the ARIMA regression model when some prior/auxiliary information is available about insignificant covariates in the model. The asymptotic properties of the shrinkage estimators, under local alternatives, are implemented including the derivations of the asymptotic distributional bias and the asymptotic quadratic risk. These results showed the effectiveness of the suggested estimation technique. Monte Carlo experiments were conducted to demonstrate the superiority of the shrinkage estimators over the MLE.
In Chapter 3, we propose the pretest and shrinkage approaches in estimating the regression parameters of the generalized autoregressive moving average (GARMA) model, which are pervasive for modelling binary and count time series data. We want to estimate regression and ARMA parameters when some of the regression parameters may belong to a subspace. We establish the asymptotic distributional biases and risks of the proposed estimators and evaluate their relative performance with respect to the unrestricted maximum partial likelihood estimator. The performance of the proposed estimators is investigated using simulation studies and real-life applications.

Chapter 4 focuses on pretest and shrinkage estimation strategies for generalized autoregressive fractionally integrated moving average (GARFIMA) models when some of the regression parameters can be restricted to a subspace. These estimation strategies are constructed on the assumption that some covariates are not statistically significant for the response. We enlighten the statistical properties of the shrinkage and pretest estimators in terms of asymptotic bias and risk. We show that the shrinkage estimators have a lower relative mean squared error as compared to the UMLE, when the number of significant covariates exceeds two. Monte Carlo simulations are conducted to examine the relative performance of the proposed estimators to the UMLE. An empirical application is used for the usefulness of our proposed estimation strategies. Finally, we summarize the findings of the thesis in Chapter 5. Also, some problems for future research are outlined in Chapter 5.
Acknowledgement

I would like to express my deepest gratitude to my thesis supervisors Professors Andrei Volodin and Shakhawat Hossain for their invaluable guidance, help, continuous encouragement, and financial support throughout my Ph.D. journey. I am grateful to them for their advice and deep involvement during my study. This dissertation would not have been completed without their relentless academic advice. My thanks extend to all faculty, staff members, and graduate students in the Department of Mathematics and Statistics. Also, I would like to acknowledge the University of Regina for financially supporting my Ph.D. study.
Post-Defence Acknowledgement

My thanks also go to the external examiner Dr. Umberto Triacca and the advisory committee members Dr. Dianliang Deng, Dr. Yiyu Yao, and Dr. Arzu Sardarli for their careful review of my thesis and for providing with useful suggestions and comments which have improved my dissertation.
Dedication

To My Family
Contents

Abstract i

Acknowledgement iii

Post-Defence Acknowledgement iv

Dedication v

Table of Contents vi

List of Tables xi

List of Figures xii

Abbreviations xiv

List of Symbols xvii

Statement of Contributions xix
1 Introduction

1.1 Introduction and Background ....................................... 1

1.1.1 Unrestricted and Restricted Estimators ...................... 3

1.1.2 Pretest Estimator .................................................. 6

1.1.3 Shrinkage and Positive Shrinkage Estimators ............... 6

1.1.4 Improved Pretest Estimator ..................................... 7

1.2 Appraisal of the Estimators ......................................... 8

1.2.1 Asymptotic Comparison ........................................... 8

1.2.2 Monte Carlo Simulation Study .................................. 10

1.3 Review of Literature .................................................. 10

1.3.1 Shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors .............................................. 10

1.3.2 Generalized Autoregressive Moving Average Models: An Efficient Estimation Approach .................................................. 12

1.3.3 Efficient estimation method for generalized ARFIMA models . 13

1.4 Objective of the Study ............................................... 14

1.5 Organization of the Thesis and Highlights of Contributions .... 15

2 Shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors

2.1 Introduction ............................................................ 17
2.2 Models and Estimation Strategy ................................................. 21
  2.2.1 Linear Regression model with ARIMA($p,d,q$) Errors ............... 21
  2.2.2 Estimation of Parameters .................................................. 22
  2.2.3 Modified Newton Raphson procedure for ML estimates .......... 25
  2.2.4 Restricted estimators ...................................................... 26
  2.2.5 Shrinkage and Positive Shrinkage Estimators ....................... 27
  2.3 Asymptotic Results ............................................................ 28
    2.3.1 Asymptotic Distributional Results: Bias ......................... 28
    2.3.2 Asymptotic Distributional Results: Risk ......................... 30
  2.4 Simulation Results ........................................................... 32

  3 Generalized Autoregressive Moving Average Models: An Efficient Estimation Approach ............................................. 38
    3.1 Introduction ................................................................. 39
    3.2 Model specification of GARMA ............................................ 44
    3.3 Estimation of GARMA models ............................................. 46
      3.3.1 Conditional Information Matrix .................................... 50
      3.3.2 The restricted maximum partial likelihood estimator .......... 53
      3.3.3 Shrinkage and improved pretest estimators ..................... 54
    3.4 Asymptotic results .......................................................... 56
    3.5 Simulation ................................................................. 62
3.5.1 ADR analysis when $\Delta = 0$ ................................. 68
3.5.2 ADR and MSE analyses when $\Delta \geq 0$ ..................... 70
3.6 Real data example: binary response .............................. 73
3.7 Concluding remarks .............................................. 78
3.8 Appendix ......................................................... 79

4 Efficient estimation method for generalized ARFIMA models 88
4.1 Introduction ...................................................... 89
4.2 Model specification of GARFIMA ................................. 93
4.3 Estimation of GARFIMA models ................................. 96
  4.3.1 Partial Score Vector ........................................ 97
  4.3.2 Conditional Information Matrix ............................ 98
  4.3.3 The restricted maximum partial likelihood estimator .... 100
  4.3.4 The pretest and shrinkage estimators ..................... 101
4.4 Asymptotic results .............................................. 104
4.5 Simulation results .............................................. 109
4.6 Real Data Application .......................................... 115
4.7 Conclusion ..................................................... 120
4.8 Appendix ......................................................... 121

5 Conclusion and Discussion ........................................... 131
5.1 Conclusion and Future Work ................................. 131

5.1.1 Future Scope ............................................. 134
List of Tables

2.1 RMSEs of $\hat{\xi}_r$, $\hat{\xi}_S$, and $\hat{\xi}_{S+}$ with respect to $\hat{\xi}$ when the restricted parameter space is correct ($\Delta = 0$)  

3.1 RMSEs of RMPLE, SE, PSE, IPT for binary and count responses  

3.2 Estimates (first row $\times 10^{-2}$) and standard errors (second row $\times 10^{-2}$) for significant covariates $PC_{t-1}(\tilde{\gamma}_1)$, $PC_{t-5}(\tilde{\gamma}_5)$, $PC_{t-6}(\tilde{\gamma}_6)$, and $PC_{t-7}(\tilde{\gamma}_7)$  

4.1 RMSEs of RMPLE, PT, SE, and PSE with respect to $\hat{\gamma}_1$ when the restricted parameter space is correct ($\Delta = 0$), and $n = 250$. 

4.2 RMSEs of RMPLE, PT, SE, and PSE with respect to $\hat{\gamma}_1$ when the restricted parameter space is correct ($\Delta = 0$), and $n = 350$. 

4.3 Estimates (first row) and standard errors (second row) for $TB_{t-1}(\tilde{\gamma}_1)$, $P_{t-4}(\tilde{\gamma}_2)$, $P_{t-5}(\tilde{\gamma}_3)$, $P_{t-6}(\tilde{\gamma}_4)$, $P_{t-7}(\tilde{\gamma}_5)$
List of Figures

2.1 RMSEs of RE(\(\hat{\xi}_r\)), SE (\(\hat{\xi}_S\)), and PSE (\(\hat{\xi}_{S+}\)) with respect to URE(\(\hat{\xi}\)) when the subspace misspecifies \(\Delta \geq 0\), \(n = 500, 600\), and \(k_1 = 3, k_2 = 8, 12, 15, 18\). ................................. 35

2.2 RMSEs of RE(\(\hat{\xi}_r\)), SE (\(\hat{\xi}_S\)), and PSE (\(\hat{\xi}_{S+}\)) with respect to URE(\(\hat{\xi}\)) when the subspace misspecifies \(\Delta \geq 0\), \(n = 700, 800\), and \(k_1 = 3, k_2 = 8, 12, 15, 18\). ................................. 36

3.1 RMSEs of RMPLE, SE, PSE, and IPT for binary and Count response models with respect to UMPLE when the subspace misspecifies \(\Delta \geq 0\), \(n = 200, 300\), and \(l_1 = 3, l_2 = 5, 15\). ................................. 66

3.2 RMSEs of RMPLE, SE, PSE, and IPT for binary and Count response models with respect to UMPLE when the subspace misspecified \(\Delta \geq 0\), \(n = 200, 300\), and \(l_1 = 5, l_2 = 3, 13\). ................................. 68
3.3 MSEs of URMPLE, RMPLLE, SE, PSE, and IPT for binary and count response models when the subspace misspecifies $\Delta \geq 0$, $n = 200, 300$, and $l_1 = 3, l_2 = 5, 15$. .......................... 70

3.4 MSEs of URMPLE, RMPLLE, SE, PSE, and IPT for binary and count response models when the subspace misspecifies $\Delta \geq 0$, , $n = 200, 300$, and $l_1 = 5, l_2 = 3, 13$ .......................... 72

3.5 Graphical displays of S&P 500 indexes from December, 1994 to February, 2021 for (a) the one-month excess stock return, (b) the S&P 500 closing price, and (c) the 3-month U.S. Treasury Bill. ............... 75

3.6 ACF and partial ACF for Binary response. .................. 76

4.1 The ACF plot based on simulated data .................. 110

4.2 RMSEs of RMPLLE, PT, SE, and PSE with respect to $\hat{\gamma}_1$ of the estimators when the subspace misspecifies $\Delta$ as zero and response is binary for $n = 250$ and $n = 350$ ............................. 113

4.3 Time series plots of (a) monthly excess stock return, (b) the S&P500 closing price, and (c) the 3-month U.S. Treasury Bill .............. 116

4.4 (a) Autocorrelation function (ACF) plot of RMPLLE model model. (b) Partial autocorrelation function (PACF) plot of the RMPLLE model for the S&P500 data. ............................. 118
Abbreviations

ACF = autocorrelation function
AIC = Akaike information criteria
ADB = asymptotic distributional bias
ADR = asymptotic distributional risk
AR = autoregressive
MA = moving average
ARMA = autoregressive moving average
ARIMA = autoregressive integrated moving average
ARFIMA = autoregressive fractionally integrated moving average
GARCH = Generalized autoregressive conditional heteroskedasticity
GARMA = Generalized autoregressive moving average
GARFIMA = Generalized autoregressive fractionally integrated moving average
GLM = generalized linear model
GLS = generalized least squares
GPLSIM = generalized partially linear single-index models
IM = information matrix
INAR = integer-autoregressive
IPT = improved pretest estimator
IRLS = iteratively reweighted least squares
LAD = least absolute deviation
LASSO = least absolute shrinkage and selection operator
ML = maximum likelihood
MSE = mean squared error
NSI = non-sample information
PACF = partial autocorrelation function
PSE = positive-part shrinkage estimator
PT = pretest estimator
QB = quadratic bias
RegARIMA = regression model with autoregressive integrated moving average
RMPLE = restricted maximum partial likelihood estimator
RMSE = relative mean squared error
SE = shrinkage estimator
SP = standard poor
UPI = uncertain prior information
URMPLE = unrestricted maximum partial likelihood estimator

VARMA = vector autoregressive moving average
List of Symbols

$\beta$ - regression parameter vector

$H_0$ - null hypothesis

$I$ - identity matrix

$l$ - the number of regression parameters

$K(n)$ - Pitman type local alternative hypothesis

$L$ - weighted quadratic loss function

$n$ - sample size

$\hat{\beta}$ - unrestricted maximum partial likelihood estimator

$\tilde{\beta}$ - restricted maximum partial likelihood estimator

$\hat{\beta}_p$ - pretest estimator

$\hat{\beta}_s$ - shrinkage estimator

$\hat{\beta}_{s+}$ - positive-part shrinkage estimator

$\hat{\beta}_{ipt}$ - improved pretest estimator

$\hat{D}_n$ - log-partial likelihood ratio test
$I(F)$ - indicator function of a set $F$

$W$ - a positive semi-definite weight matrix in the quadratic loss function

$\Delta$ - non-centrality parameter

$\Psi_{\nu}(x, \Delta)$ - noncentral $\chi^2$ distribution function with non-centrality parameter $\Delta$

and degrees of freedom $\nu$

$G_n(\beta)$ - Conditional Information matrix

$\Sigma$ - variance covariance matrix

$d$ - a fixed real valued vector contained in $K(n)$. 
Statement of Contributions

I. Co-Authorship Declaration

I hereby declare that this thesis incorporates the outcome of joint research undertaken in collaboration with my supervisors, Professor Andrei Volodin and Professor Shakhawat Hossain. In all cases, the key ideas, primary contributions, experimental designs, data analysis, and interpretations, were performed by the author, and the contribution of the co-author was primarily through the provision of some theoretical results.

I am aware of the University of Regina Senate Policy on Authorship, and I certify that I have properly acknowledged the contribution of other researchers to my thesis and have obtained written permission from each of the co-authors to include in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers is the product of my own work.
II. Declaration of Previous Publication

This thesis includes two original papers that have been previously published and another one is submitted for publication:

<table>
<thead>
<tr>
<th>Thesis Chapter</th>
<th>Publication title/ full citation</th>
<th>Publication status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 2</td>
<td>Shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors.</td>
<td>Submitted</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Generalized Autoregressive Moving Average Models: An Efficient Estimation Approach</td>
<td>Published</td>
</tr>
<tr>
<td></td>
<td>Journal of Statistical Computation and Simulation</td>
<td></td>
</tr>
<tr>
<td>Chapter 4</td>
<td>Efficient estimation method for generalized ARFIMA models. Communications in Statistics - Theory and Methods.</td>
<td>Published</td>
</tr>
</tbody>
</table>

I certify that I have obtained written permission from the copyright owner to include the above-published material in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Regina. I declare that, to the best of my knowledge, my thesis does not infringe upon anyone’s copyright nor violate any proprietary rights and that any ideas,
techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained written permission from the copyright owner to include such material in my thesis. I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office and that this thesis has not been submitted for a higher degree to any other university.
Chapter 1

Introduction

1.1 Introduction and Background

Despite many papers published on time series models, there have been moderate investigation on how to predict numerical values that depend on the whole series, rather than depending more on recent than past values. The word regression has many different meanings in different contexts. In the machine learning context, it means predicting a numerical value from a set of predictors. With respect to forecasting time series, regression typically means fitting the historical time series data with a regression model such as ARIMA (Box and Jenkins, 1970), GARMA Benjamin et al. (2003)[9]) and ARFIMA Tsai (2009) [66] models to forecast future values of the time series. These time series models heavily depend on recent or seasonal values. These models may be useful when a linear model is not appropriate and this situation can often occur in practice in many applications. Examples include price fluctuations on
financial markets, sensor readings in industrial processes (e.g. the temperature or pressure in a chemical reactor), patients’ medical records (e.g. the blood pressure and heart rate), and customers’ shopping histories.

Time series regression models are used to obtain information about unknown parameters. Estimating the unknown parameters of unrestricted model is important for conducting the statistical tests on any single or set of model parameters. However, the classical estimators of the unknown parameters of time series regression models are based on the sample information. In real life situations, researchers may have some prior/auxilliary information on the parameters either in the form of a prior distribution or as a constraint on some (all) of the parameters. This information may be considered as non-sample information (NSI). The NSI may or may not contribute in a positive manner in the estimation process. However, it could be beneficial to use of NSI in the estimation procedure when sample-information may be limited and may not be entirely trustworthy. Information from the NSI or from the model selection procedure are incorporated in the estimation procedure to lead to restricted models. The performance of the restricted models compared to the unrestricted models depends on the quality of NSI.

In order to make determined efforts to deal with the NSI in the estimation process, it is natural to perform a pretest or preliminary test on the validity of NSI. This NSI is then either used or not depending upon the outcome of the statistical test. The
resulting estimation procedure is called a pretest or preliminary test estimator and this estimator was proposed by Bancroft (1944)[7]. Later in 1956, Stein provided shrinkage estimators that combine the sample and NSI in such a way that helps to improve the accuracy of the estimation process. These estimators can be obtained by shrinking unrestricted model estimator in the direction of the restricted model estimator. The superiority of this estimator can be seen in terms of risk in this dissertation with respect to the maximum partial likelihood estimators for the case of a few time series models. Sclove (1972)[60] introduced improved pretest estimators that perform better than pretest estimators. This dissertation deals with a pretest, improved pretest, and shrinkage estimation strategies for estimating the parameters of time series regression models, such as multiple linear regression with ARIMA errors, generalized ARMA(GARMA), and generalized ARFIMA(GARFIMA) models.

1.1.1 Unrestricted and Restricted Estimators

In the time series regression model, an estimator considers the sample information as a given set of covariates to predict the response variable. During the prediction of the response variable, it is conjecture that there is a subset of the covariates in the sample information which does not contribute significantly. In some circumstances, sample information containing a subset of the covariates may be considered as a nuisance such that they are not of main interest, but they must be taken into account
in estimating the remaining parameters. When an estimator is solely based on sample
information and not a function of NSI, it is called the unrestricted estimator (UE).
Denote the UE of $\beta$ by $\hat{\beta}$. This estimator can be achieved through different methods,
such as generalized least square (GLS), maximum likelihood (ML), maximum partial
likelihood (MPL), etc.

On the other hand, if estimation methods are a function of NSI in addition to the
sample information, it would result in a better estimator than using sample informa-
tion alone in the estimation process. A function of NSI can be written in the form of
a general linear constraint as:

$$H_0 : R\beta = h,$$

where $R$ is $l_2 \times l$ matrix of rank $l_2$ and $h$ is a $l_2 \times 1$ vector of constants and $\beta$ is a $l \times 1$
vector of model parameters. A model with restrictions on parameters given in (1.1)
is a reduced model or sub-model while a model with no restriction on parameters is
called a full model.

As a particular form of the above general linear constraint (null hypothesis) in
(1.1), consider the case $R = [O, I]$, where $O$ and $I$ are zeros and identity matrices
of orders $l_2 \times l_1$ and $l_2 \times l_2$ respectively. In this case, the parameter vector $\beta$ can be
partitioned as $\beta = (\beta_1^T, \beta_2^T)^T$ with order $l_1 \times 1$ of $\beta_1$ and $l_2 \times 1$ of $\beta_2$ respectively,
such that $l = l_1 + l_2$ where $\beta_1$ is the vector of main effects and $\beta_2$ relates to the
nuisance parameters which can be excluded from the model. The general form (1.1)
can be reduced as:

\[ H_0 : \beta_2 = 0, \quad (1.2) \]

with \( l_1 \) and \( l_2 \) dimensions of \( \beta_1 \) and \( \beta_2 \) such that \( l = l_1 + l_2 \). Therefore, in order to take the advantage of the available NSI in the estimation process, we define \( \tilde{\beta} \), the restricted estimator (RE) of \( \beta \). When NSI holds, \( \tilde{\beta} \) will be an unbiased estimator of \( \beta \) with a smaller variance than \( \hat{\beta} \). However, if the NSI is not true, then \( \tilde{\beta} \) is a biased estimator, and the \( \hat{\beta} \) outperforms \( \tilde{\beta} \).

Here, a question arises as to how one incorporates this NSI into the estimation procedure of model parameters. This dissertation deals with an improved estimation strategy of parameter in the time series regression models. This strategy is known as the James-Stein estimation strategy, also called the shrinkage estimation, inspired by Stein’s result that shows in a parameter dimension greater than two. Efficient estimates can be obtained by shrinking full model estimates in the direction of reduced model estimates. We consider the pretest, improved pretest, shrinkage, and positive-part shrinkage estimators to three different time series regression models where sample information, as well as NSI about the model parameters, are available. We discuss these proposed estimators in the following subsections.
1.1.2 Pretest Estimator

Let $D_n$ be the test statistic for testing the null hypothesis in (1.1) and $\chi^2_{l_2, \alpha}$ is the $\alpha$-level critical value of the chi-square distribution of $D_n$ under $H_0$ with $l_2$ degrees of freedom. Bancroft (1944)[7] introduced the pretest estimator (PT) of $\beta$ based on $\hat{\beta}$ and $\tilde{\beta}$ and is defined as follows:

$$
\hat{\beta}_p = \hat{\beta} - I(D_n \leq \chi^2_{l_2, \alpha})(\hat{\beta} - \tilde{\beta}), \quad l_2 \geq 1,
$$

where $I(F)$ is an indicator function of a set $F$. Based on the indicator function, $\hat{\beta}_p$ chooses $\tilde{\beta}$ or $\hat{\beta}$ according to whether $H_0$ is accepted or rejected, respectively. It is important to note that $\hat{\beta}_p$ is bounded and performs better than $\tilde{\beta}$ in some part of the parameter space. More useful discussions about this estimator can be found in Judge (1978)[35], Nkurunziza (2013)[43] and Ahmed(2006)[2].

1.1.3 Shrinkage and Positive Shrinkage Estimators

The pretest estimator has certain limitations, as it is a discontinuous function of $\hat{\beta}$ and $\tilde{\beta}$ and is dependent on the choice of $\alpha$ through the critical value $\chi^2_{l_2, \alpha}$. We can overcome this limitation using the shrinkage estimator (SE), which incorporates the information from both unrestricted and restricted estimators, defined as:

$$
\hat{\beta}_s = \tilde{\beta} + (1 - (l_2 - 2)D_n^{-1})(\hat{\beta} - \tilde{\beta}), \quad l_2 \geq 3.
$$
The SE is the linear combination $\lambda \hat{\beta} + (1 - \lambda) \tilde{\beta}$, where $\lambda \in [0, 1)$ is the parameter that determines the extent to which these estimates are pooled together. If $\lambda = 0$, then the restricted estimator dominates completely, whereas for $\lambda = 1$, then no shrinkage occurs. A setback with the SE is that it is not a convex combination of the unrestricted and restricted estimators. This has the tendency to cause over-shrinking which gives the estimator the opposite sign of the unrestricted estimator. To overcome this issue, we use the positive-part shrinkage estimator (PSE):

$$\hat{\beta}_{s+} = \tilde{\beta} + (1 - (l_2 - 2)D_n^{-1})^+(\hat{\beta} - \tilde{\beta}), \quad l_2 \geq 3,$$

and where $z^+ = \max(0, z)$.

### 1.1.4 Improved Pretest Estimator

If we replace $\hat{\beta}$ by $\hat{\beta}_s$ in the pretest estimator, the resulting estimator dominates the usual pretest estimator. This estimator is called the improved pretest estimator (ipt) (Sclove (1972)[60]) and it is defined by:

$$\hat{\beta}_{ipt} = \tilde{\beta} + (1 - (l_2 - 2)D_n^{-1})(\hat{\beta} - \tilde{\beta})I(D_n > \chi^2_{l_2,\alpha}), \quad l_2 \geq 3.$$

The improved pretest estimator dominates the pretest estimator over a range of parameter values even though we have the limitation $l_2 \geq 3$. If $\chi^2_{l_2,\alpha} \leq (l_2 - 2)$, $\beta_{ipt}$ behaves like $\hat{\beta}_{s+}$. For $\chi^2_{l_2,\alpha}$ outside this range, it works like $\hat{\beta}_p$ estimator but it still continues to perform better than the pretest estimator (Ahmed(1999)[1]).
1.2 Appraisal of the Estimators

1.2.1 Asymptotic Comparison

We study the efficiency of $\hat{\beta}$, $\tilde{\beta}$, $\hat{\beta}_p$, $\hat{\beta}_s$, $\hat{\beta}_{s+}$ and $\hat{\beta}_{ipt}$ using the notion of asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR). In general, it is not an easy task to achieve the finite sample properties of the shrinkage and pretest estimators for non-normal models. Asymptotic methods have overcome this difficulty (Hossain(2016)[32]) which is related to convergence in distribution but does not guarantee convergence in quadratic risk. By implementing the notion of asymptotic distributional risk (ADR), this technicality will be taken care of and it plays a useful role in asymptotic risk analysis. For this aim, we consider the weighted quadratic loss function criterion to examine the performance of the estimators:

$$L(\beta^*; W) = (\sqrt{n}(\beta^* - \beta))^T W (\sqrt{n}(\beta^* - \beta)),$$

where $\beta^*$ is any of the proposed estimators $\hat{\beta}$, $\tilde{\beta}$, $\hat{\beta}_s$, $\hat{\beta}_{ipt}$, or $\hat{\beta}_{s+}$ and $W$ is a positive semi-definite weight matrix. Here, we recommend $N = I$ as this provides the quadratic loss function and give the guarantee that the shrinkage estimators perform better than the UMPLE estimator as compared to other weight matrices. The expectation of the loss function:

$$\text{ADR}(\beta^*; W) = \mathbb{E} \left( \lim_{n \to \infty} L(\beta^*; W) \right)$$

is called the asymptotic distributional risk.
We can evaluate the performance of the estimators by comparing the ADR with a suitable matrix $W$. The smaller the ADR, the better the estimator. If there exists another estimator $\beta^0$, such that $\text{ADR}(\beta^0; W) \leq \text{ADR}(\beta^*; W)$ for all $(\beta, W)$ with strict inequality for some $\beta$, then the estimator $\beta^*$ will be called an inadmissible estimator. In such cases, we say that the estimator $\beta^0$ dominates $\beta^*$.

Ahmed (1997) noted that since the statistic $D_n$ is consistent against the fixed alternative, the SE and PSE will be asymptotically equivalent in probability to UE, (i.e., the asymptotic distribution of $\sqrt{n}(\beta^* - \beta)$, is equivalent to $\sqrt{n}(\hat{\beta} - \beta)$ as $n \to \infty$). Therefore, for large sample situations, there is not much to investigate on the estimators. In this case, to obtain meaningful asymptotic results and to evaluate the behavior of the estimators in a neighborhood of the null hypothesis, a class of local alternatives, $K(n)$, is considered, which is given by:

$$K(n) : \mathbf{R}\beta = \delta \sqrt{n},$$

where $\delta = (\delta_1, \delta_2, \ldots, \delta_{l_2}) \in \mathbb{R}^{l_2}$ is a fixed vector of real numbers with size $l_2 \times 1$.

There is no doubt $\delta = 0$ means that $\mathbf{R}\beta = 0$, which is a particular case of (4.9). The ADB of the estimator $\beta^*$ is defined as:

$$\text{ADB}(\beta^*) = \mathbb{E}\left(\lim_{n \to \infty} \sqrt{n}(\beta^* - \beta)\right).$$

Now, we define the asymptotic distributional risk (ADR) of $\beta^*$ by:

$$\text{ADR}(\beta^*; W) = \int \cdots \int z^\top W z dG(z) = \text{tr} \left[ W \Sigma^*(\beta^*) \right],$$
where $\Sigma^*(\beta^*) = \lim_{n \to \infty} E \left( \sqrt{n} (\beta^* - \beta) \right) = \int \cdots \int z^T dG(z)$ is the dispersion matrix for the distribution $G(z)$. We then compare the behavior of the estimators based on the ADB and ADR.

1.2.2 Monte Carlo Simulation Study

Along with the bias and risk comparisons of the proposed estimators, we also carry out Monte Carlo simulation studies to investigate the comparison of $\hat{\beta}, \hat{\beta}_p, \hat{\beta}_s, \hat{\beta}_{s+}$ and $\hat{\beta}_{ipt}$ with respect to the $\hat{\beta}$.

1.3 Review of Literature

In the following subsections we give an introduction and literature review of three models that were used in this dissertation.

1.3.1 Shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors

In chapter 2, we consider the shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors. In the literature, there are many papers that consider the James-Stein shrinkage method for parameter estimation in the linear regression model with known autocorrelated error, AR, ARMA, and GARCH errors.
Wang (2007)[68] introduced a modified least absolute shrinkage and selection operator (lasso) in the regression model with autoregressive errors and its superiority compared with a traditional lasso in the context of the Bayes information criterion through a simulation study and a practical example of electricity demand data set. Saleh (2009)[59] introduced improved estimation of the parameters of a simple linear regression model with known autocorrelated errors. Results showed that the shrinkage estimator and the preliminary test estimator dominated the unrestricted estimator in the context of bias and MSE. Alheety (2011)[4] worked on the ridge-type shrinkage estimators in the linear regression with AR error. The simulation study suggests that some ridge-type shrinkage estimators are promising and can be recommenced for practitioners.

Wu (2012)[73] proposed a shrinkage procedure for linear regression model with ARMA error which simultaneously estimates the parameters and selects the informative variables in the regression, autoregressive, and moving average components. Also, an empirical application of ground-level ozone time series is considered with a simulation study. Hossain (2014)[30] developed shrinkage estimation for linear regression model with AR errors. In addition, the authors studied the asymptotic features of the estimators in the context of risks and biases. The utility of the proposed estimators was applied to Los Angeles pollution mortality data. Hossain (2016)[32] considered shrinkage and positive shrinkage estimators in linear regression with GARCH error.
We extend this work in linear regression with ARIMA error in this chapter. Paolella (2019)[52] worked on the application of a shrinkage estimator for linear models with ARMA and GARCH errors.

1.3.2 Generalized Autoregressive Moving Average Models: An Efficient Estimation Approach

Categorical and count time series data have been attracting considerable attention both in terms of data analysis and in the development of methodological approaches. In this regard, the use of classical models has a disadvantage in providing an exact parametrization of conditional distributions with the help of normal distribution. Benjamin et al. (2003) [9] overcome this problem by introducing GARMA models that extend the regression model with ARMA error to an amenable data-driven model for non-normal time series with the help of the exponential family.

In the literature of the GARMA model, Woodard (2011)[72] presented some general results on stationary and ergodic solutions for some observation-driven GARMA models for the binary, count, or other discrete-valued data which are similar to those available in Benjamin et al. (2003)[9]. An observation-driven ARMA model for count time series data with covariates implemented in the glarma R package (Dunsmuir (2015)[16]). Albarracin (2019)[3] focused on GARMA models where the conditional
mean depends on covariates and also lagged values of the response in the set of regressors to model the serial dependence. They made an inference about the effects of covariates on the mean response and also took into account serial dependence. Moller (2020)[41] proposed a generalized discrete ARMA approach for stationary multivariate or univariate time series data of integers, counts, and compositional data. Hossain (2016)[32] studied time series following GLM and applied the pretest and shrinkage techniques to estimate the mean response parameters.

1.3.3 Efficient estimation method for generalized ARFIMA models

The GARMA model is useful to capture short-range dependence for the time series data set. A drawback of GARMA models may occur in the case of non-stationarity in long-memory high-frequency time series data. The main issue of the GARMA model is covered by GARMFIMA as this model may play an important role in modeling long-memory high-frequency time series data in the field of economics, medicine, biology, and hydrology.

Therefore, in Chapter 4, we extend the pretest and shrinkage estimation method of GARMA models to GARMFIMA models. Basically, GARMFIMA is a generalization of the GARMA\((p,q)\) model. In the literature, Reisen (2001)[56] conducted extensive numerical studies to evaluate the procedures for estimating the parameters of an
ARFIMA process for both parametric and semiparametric methods and the results indicate that the regression methods outperforms the parametric Whittle’s method when AR or MA components are involved. Doornik (2003)[15] implemented the improvement of computational properties of ARFIMA models by computing the quasi-likelihood exactly, using the autocovariance functions up to order $n$. Palma (2007)[49] was implemented statistical tools for analyzing ARFIMA in terms of predictive capability, exact autocovariance, and parameter estimation. Rocha (2009)[57] proposed $\beta$ARMA model that includes both autoregressive and moving average dynamics and also includes a set of regressors. Tsai (2009) [66] compared the continuous-time ARFIMA model with discrete-time ARFIMA for examining time series data of short and long memory dependence. Pumi (2018)[54] introduced a class of beta ARFIMA models for continuous random variables taking values in the continuous unit interval $(0, 1)$. They accommodated a set of regressors and a long-range dependent time series structure in the model.

1.4 Objective of the Study

The objectives of this dissertation are as follows:

i. To implement the pretest, improved pretest, shrinkage, and positive shrinkage estimators for some time series regression models.
ii. To compare the relative performance of the estimators theoretically and numerically, with respect to the classical estimators.

iii. To apply the above estimators to real data sets.

1.5 Organization of the Thesis and Highlights of Contributions

The organization of the thesis is as follows:

i. Chapter 1 presents the introduction, review of the literature, objectives of the study, and important findings.

ii. Chapter 2 discusses the restricted, shrinkage, and positive-part shrinkage estimators in linear regression model with ARIMA error. The asymptotic distribution, asymptotic biases, and risks of the estimators will also be derived. The performance of the estimators is investigated by extensive simulation studies.

iii. In Chapter 3 we propose shrinkage, positive shrinkage, and improved pretest estimators for the GARMA model. Theoretical results on the risks and biases of the unrestricted, restricted, improved pretest, shrinkage, and positive shrinkage estimators are derived based on the concept of distributional biases and risks. Also, mean squared error matrices of these estimators are derived
and compared analytically using MLE as a benchmark estimator for the comparison. We study numerically, their relative performances with respect to the unrestricted maximum likelihood estimator. Numerical analyses are conducted via Monte Carlo simulation. We apply these estimation procedures to S&P500 stock market data. In this application to a real data set, we device a bootstrapping procedure to obtain the relative mean squared errors of the proposed estimators.

iv. Chapter 4 extends the proposed estimators of Chapter 3 for the long-range dependence time series structure with the GARFIMA model. We propose the restricted, pretest, and shrinkage estimators for estimating the regression parameters and autoregressive and moving average parameters. These estimators are new for ARFIMA models and have never been considered in the literature. Following the structure of Chapter 3, we study numerically their relative performances. We further illustrate the proposed methodology through an analysis of a popular econometric data set.

v. Chapter 5 presents our conclusions of this thesis and a brief discussion of future works.
Chapter 2

Shrinkage estimation of a linear regression model with ARIMA (RegARIMA) errors

2.1 Introduction

In the time series applications of the linear regression model, significant serial dependence might exist between response and predictor variables. In such circumstances, the error terms may be in the form of a non-stationary serially dependent process in the model. Hence, the linear regression model with autoregressive integrated moving average (RegARIMA) errors is of particular interest, as in many situations the structure of the dependence can be accommodated by choosing the appropriately ARIMA process.

In the last five decades, the RegARIMA models and the linear regression model with autoregressive moving average errors (RegARMA) constantly appear for the
modeling of time series data. For Instance, Alupotha (2021)[5] applied the RegARIMA model to study the government expenditure in Sri Lanka in an annual authentic data set from 1981 to 2016, and standard statistical techniques were employed for model verification in terms of mean absolute percentage error.

Johansen (2012)[34] introduced $C_p$ statistics for regression models with stationary and non-stationary ARIMA errors, in which asymptotic features of maximum likelihood estimators with simulation study were discussed in detail. An outstanding example of linear regression models with ARIMA component errors in the Fortran language is presented by William et al. (2011)[70], discussing computational and theoretical outcomes of Gaussian maximum likelihood (ML) estimation using three real-life applications of time series. Bianco (2001)[10] used a robust estimation approach to detect the outliers in a regression model with ARIMA errors, and simulation results were compared with classical methods based on maximum likelihood type estimates and Kalman filtering. Davis (1997)[14] employed the least absolute deviation (LAD) estimations method in a linear regression model with ARMA errors under the general conditions. The asymptotic properties of the estimator for the LAD objective function rely on the functional limit theorem. William (1987)[69] considered an iterative general least square (GLS) approach to maximum likelihood estimation of regression models with ARIMA Errors. Wince (1986)[71] considered the exact maximum likelihood estimation of the RegARMA model with possibly consecutive and
nonconsecutive time series data. In this paper, we consider the James-Stein shrinkage estimation procedure for the RegARIMA under the availability of prior/auxilliary information about insignificant covariates in the model. Any statistical model, and even RegARIMA, may lose forecasting capability due to the countless selection of insignificant covariates in the model. In this situation, the best estimation strategy is to use the shrinkage procedure that estimates the coefficients of significant covariates by utilizing the knowledge included in discarded insignificant covariates. Therefore, we can split the all parameter vector, $\xi$, in the form of two sub-vectors, $\xi = (\xi_1^T, \xi_2^T)^T$, where $\xi_1$ and $\xi_2$ are assumed to have dimensions $(k_1 + p + q) \times 1$ and $k_2 \times 1$, respectively, such that $k = k_1 + p + q + k_2$. Our goal in this paper is to estimate parameter sub-vector $\xi_1$ utilizing the information of $\xi_2$ during the estimation process. We set the values of $\xi_2$ near some specified values along with the hypothesis $H_0 : \xi_2 = 0$. The general form of hypothesis can be expressed as $H_0 : R\xi = 0$, where $R = [I_{k_2}, 0_{k_2 \times (k-k_2)}]$ and $0_{a \times b}$ is an $a \times b$ matrix of zeros. This parameter partition strategy was used by Hossain (2016)[32] for a linear regression model with GARCH errors and employed shrinkage techniques to estimate the mean response parameters. In this work, we expand this methodology for the RegARIMA model.

In recent literature, researchers have shown significant interest in the usefulness of James-Stein shrinkage method for parameter estimation in the linear regression model.
with AR, ARMA, and GARCH errors. Paolella (2019)[52] worked on the application of shrinkage estimator for linear models with ARMA and GARCH errors. Hossain (2016)[32] considered shrinkage and positive shrinkage estimators in linear regression with GARCH error. Thomson (2015)[64] developed shrinkage estimation in a linear regression model with AR errors. In addition, the authors studied the asymptotic features of the estimators in the context of risks and biases. The utility of the proposed estimators was tested on Los Angles pollution mortality data. Wu (2012)[73] proposed a shrinkage procedure for linear regression model with ARMA, error which simultaneously estimates the parameters and selects the informative variables in the regression, autoregressive, and moving average components. Also, an empirical application of ground-level ozone time series was considered with a simulation study. Wang (2007)[68] introduced a modified least absolute shrinkage and selection operator (lasso) in the regression model with autoregressive errors, and its superiority was compared with a traditional lasso, in the context of the Bayes information criterion, through a simulation study and a practical example of electricity demand data set.

The paper is structured as follows. In Section 2.2, we present the model specification, parameter estimation, score vector, and Hessian Matrix of RegARIMA model. Section 2.3 reviews the asymptotic distributional bias and risk of the proposed estimators. In Section 2.4, we consider a Monte Carlo simulation study to evaluate the numerical performance of the exact maximum likelihood estimator. An application
of RegARIMA models is represented in Section 2.5. Section 2.6 closes the article with concluding remarks.

2.2 Models and Estimation Strategy

2.2.1 Linear Regression model with ARIMA\((p, d, q)\) Errors

We assume a multiple linear regression model as:

\[
y_t = x_t^T \beta + \eta_t, \quad t = 1, 2, \ldots, N = n + d,
\]

where \(y_t\) is the response, \(x_t = (x_{t1}, x_{t2}, \ldots, x_{tk})^T\) is a \(k \times 1\) predictor vector, and \(\beta = (\beta_1, \beta_2, \ldots, \beta_k)^T\) is a \(k \times 1\) vector of unknown regression coefficients.

We assume that the error term \(\eta_t\) in (2.1) is generated by an ARIMA model of order\((p, d, q)\), where \(p\) is the order of autoregressive part, \(d\) is the degree of differences, and \(q\) is the order of moving average part. From \(\eta_t\), we can modify error term \(\omega_t\) of \((n + d) - d\) differences \(\omega_1, \omega_2, \ldots, \omega_n\), where \(\omega_t = \nabla^d \eta_t\). While the differences appear in \(\eta_t\) of (2.1), all corresponding series (both of the dependent and the explanatory variables) should also occur the difference Pankratz (1991)[51]. Thus, we applied \((n + d) - d\) differences on \(y_t\) and \(x_t\). So the multiple linear regression model in (2.1) can be re-written as:

\[
y_t^* = x_t^* \beta + w_t, \quad t = 1, 2, \ldots, n,
\]
where \( y_t^* = \nabla^d y_t \) and \( x_t^{*\top} = \nabla^d x_t^{\top} \). George (2008)[24] mentioned that the general problem of fitting the parameters \( \phi = (\phi_1, \phi_2, \ldots, \phi_p)^\top \) and \( \theta = (\theta_1, \theta_2, \ldots, \theta_q)^\top \) of the ARIMA error \( (\eta_t) \) is equivalent to fitting \( \omega_t \) as stationary ARMA\((p,q)\) error which can be written as:

\[
\omega_t = \phi_1 \omega_{t-1} + \phi_2 \omega_{t-2} + \cdots + \phi_p \omega_{t-p} + a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_q a_{t-q} \quad (2.3)
\]

with \( \omega_t = \nabla^d \eta_t \), and \( a_t \) will be referred to as generating innovations, are assumed to be independent random variables with mean 0 and common variance \( \sigma^2 \). Hence, linear regression with ARIMA\((p,d,q)\) errors in (2.1) is equivalent to linear regression with stationary ARMA\((p,q)\) errors in (2.2).

### 2.2.2 Estimation of Parameters

In the past, researchers studied the linear regression models with ARMA and ARIMA errors based on the maximum likelihood estimation approach. For instance, Furno (1996)[23] proposed an information matrix (IM) approach to study the linear regression model with ARMA errors by maximizing the likelihood function with respect to regression parameters \( \beta \) and unknown variance parameters \( \eta = (\phi^\top, \theta^\top, \sigma^2)^\top \). Wincek (1986)[71] introduced the exact maximum likelihood method for estimating the parameters for the regression model with ARMA errors. We use the same method for estimating the parameters. In a matrix form, the model (2.2) can be written in
matrix form:

\[ Y = X\beta + W \]  \hspace{1cm} (2.4)

where \( Y = (y_1^*, \ldots, y_n^*)^\top \), \( X = (x_1^*, \ldots, x_n^*)^\top \), and \( W = (w_1, \ldots, w_n)^\top \) and the vector \( W \) has a zero mean and covariance matrix \( \Gamma \) with \((t, t')th \) element \( \text{cov}(w_t, w_{t'}) = \gamma(t - t') \), where the autocovariances are functions of unknown parameters \( \eta \). We want to estimate the regression parameter vector \( \xi \) by maximizing the Gaussian log-likelihood function (2.5). The log-likelihood function of (2.4) can be written as

\[ l(\xi) = -\frac{1}{2} \log |\Gamma| - \frac{1}{2} W^\top \Gamma^{-1} W, \]  \hspace{1cm} (2.5)

where \( W = Y - X\beta \), \( \Gamma \) is a covariance matrix of \( W \) and \( \xi \) be the \((k+p+q+1) \times 1 \) vector of unknown parameters. The log-likelihood function (2.5) is a complicated function of the unknown parameters. To maximize this function, the iterative Newton-Raphson method is needed and require evaluation of the score vector and Hessian matrix of log-likelihood. An innovation transformation of \( \Gamma \) allows us to calculate the exact log-likelihood and its derivative through a convenient recursive procedure.

Wincek (1986)[71] used the innovation transformation of \( \Gamma \) to evaluate the score vector and Hessian matrix of log-likelihood \( l(\xi) \). For a covariance matrix \( \Gamma \), there exists a unique lower triangular matrix \( P \) with ones on the diagonal such that \( P^\top \Gamma P = D \), where \( D \) is a diagonal matrix of positive values. The specific prediction errors denoted by \( a_{t|t-1} = w_t - w_{t|t-1} \), are called the innovations of the \( w_t \).
Hence, the vector form of innovations $a = (a_{t|0}, a_{t|1}, \ldots, a_{t|t-1})^\top$ and $W$ are linearly related with $a = P^\top W$. Thus the second term of (2.5) can be written as $W^\top \Gamma^{-1} W = a^\top V^{-1} a / \sigma^2$, where $\Gamma = \sigma^2 V$, $V = \text{diag}(v)$, $v = (v_1, v_2, \ldots, v_t, \ldots, v_n)^\top$, $v_t = \sigma_t^2 / \sigma^2$, with $v_t$ are functions of $\eta$. The log-likelihood in (2.5) can be expressed in innovations form as

$$l(\xi) = -\frac{1}{2} n \log(\sigma^2) - \frac{1}{2} \log|V| - (2\sigma^2)^{-1} a^\top V^{-1} a,$$

where $a = P^\top(Y - X\beta)$.

The components of score vector of (2.6) are $\frac{\partial l(\xi)}{\partial \beta} = \sigma^{-2} U^\top V^{-1} a$ and $\frac{\partial l(\xi)}{\partial \eta} = F^\top N^{-1} g$, where $U = P^\top X$, $g = [a^\top (a^\top a - \sigma^2 v)^\top]^\top$, $F^\top = \left[ -\frac{\partial}{\partial \eta} a^\top \frac{\partial}{\partial \eta} (\sigma^2 v^\top) \right]$, and $N = \text{diag}(\sigma^2 V, 2\sigma^4 V^\top V)$.

The Hessian matrix of from the log-likelihood (2.6) is

$$\frac{\partial^2 l(\xi)}{\partial \xi \partial \xi^\top} = \begin{bmatrix} \frac{\partial^2 l(\xi)}{\partial \beta \beta^\top} & \frac{\partial^2 l(\xi)}{\partial \beta \eta^\top} \\ \frac{\partial^2 l(\xi)}{\partial \eta \beta^\top} & \frac{\partial^2 l(\xi)}{\partial \eta \eta^\top} \end{bmatrix} = \begin{bmatrix} \frac{\partial^2 l(\xi)}{\partial \beta \beta^\top} & 0 \\ 0 & \frac{\partial^2 l(\xi)}{\partial \eta \eta^\top} \end{bmatrix}, \quad (2.7)$$

where $\frac{\partial^2 l(\xi)}{\partial \beta \beta^\top} = -\sigma^{-2} U^\top V^{-1} U$, $\frac{\partial^2 l(\xi)}{\partial \beta \eta^\top} = -F^\top N^{-1} F$, $\frac{\partial^2 l(\xi)}{\partial \eta \beta^\top} = 0$, and $\frac{\partial^2 l(\xi)}{\partial \eta \eta^\top} = 0$. 
2.2.3 Modified Newton Raphson procedure for ML estimates

Based on Wincek (1986)[71], the approximate Hessian matrix in the Newton-Raphson procedure yields the modified Newton-Raphson equations:

\[
\hat{\beta} = \bar{\beta} + \left[ \tilde{V}^{-1} \tilde{U} \right]^{-1} \tilde{U}^T \tilde{V}^{-1} \tilde{a} = \left[ \tilde{U}^T \tilde{V}^{-1} \tilde{U} \right]^{-1} \tilde{U}^T \tilde{V}^{-1} \tilde{P}^T \tilde{Y} \tag{2.8}
\]

\[
\hat{\eta} = \bar{\eta} + \left[ \tilde{F}^T \tilde{N}^{-1} \tilde{F} \right]^{-1} \tilde{F}^T \tilde{N}^{-1} \tilde{g}, \tag{2.9}
\]

where \( \tilde{F}, \tilde{V}, \tilde{a}, \tilde{U}, \tilde{g}, \tilde{P}, \tilde{N} \) are evaluated at the \( \xi = (\tilde{\beta}^T, \tilde{\eta}^T)^T \) of the parameter vector at the previous iteration of modified Newton-Raphson procedure. The estimator \( \hat{\beta} \) in (2.8) is the weighted least squares estimate of \( \beta \) of model (2.6) and is also called the unrestricted maximum likelihood estimator (UE). This estimator \( \hat{\beta} \) expressed in terms of the transformed variables \( \tilde{U} = \tilde{P}^T X \) and \( \tilde{P}^T \tilde{Y} \) with \( \tilde{V}^{-1} \) as the diagonal matrix of weights. A suitable estimation procedure is to obtain \( \hat{\beta} \) in (2.8) and then use \( \tilde{a} = \tilde{P}^T (Y - X \tilde{\beta}) \) and \( \partial \tilde{a} / \partial \eta = (\partial \tilde{P}^T / \partial \eta)|_{\tilde{\eta}} (Y - X \tilde{\beta}) \) in (2.9) when obtaining \( \hat{\eta} \). Similar to (2.8), equation (2.9) for the adjustment \( \hat{\eta} - \bar{\eta} \) in the time series parameters also has the form of a weighted least squares estimator. Although we obtained the estimate \( \hat{\eta} \), we consider it as a nuisance parameter and primarily focus on \( \hat{\beta} \).

The above weighted least squares procedure described by (2.8) and (2.9) requires the calculation of \( a = P^T W, U = P^T X, \) and \( P^T Y, \) as well as the derivatives of \( \partial a / \partial \eta = (\partial P^T / \partial \eta) W \) and \( \partial v / \partial \eta \). To obtain these quantities with recursive procedure, see details in Wincek (1986)[71].
2.2.4 Restricted estimators

Sometimes, we are interested in estimating the regression parameters in (2.2) when some of the regression parameters may be linearly related to each other. Instead of deleting the insignificant covariate parameters, Hossain (2018)[33] incorporated the information of insignificant covariates as auxiliary information in terms of linear restrictions in the estimation process to obtain improved estimator.

We can define the hypothesis as:

\[ H_0 : \xi_2 = 0 \quad \text{vs.} \quad H_a : \xi_2 \neq 0, \quad (2.10) \]

Researchers typically test the above hypothesis (2.10), as they are believed to be a reasonable reduction of the unrestricted model. Using these restrictions, we can construct a modified log-likelihood that is, maximizing \( l(\xi) \) subject to \( R\beta = h \) is equivalent to finding

\[
\hat{\xi}_r = (\hat{\beta}_r^T, \hat{\eta}_r^T)^T = \arg\max_{\beta, \phi, \theta, \sigma^2} \{ l(\xi) : \xi_2 = 0 \}, \quad (2.11)
\]

where \( \hat{\xi}_r \) is the restricted maximum likelihood estimator(RE). The objective function (2.11) can be maximized by using the Newton Raphson method discussed in the previous section.

For testing the particular hypothesis, \( H_0 : \xi_2 = 0 \), we need the following partition.
of the expected information matrix:

\[
\mathcal{I}(\xi) = E \begin{bmatrix}
\frac{\partial^2 l(\xi)}{\partial \xi_i \partial \xi_j} & \frac{\partial^2 l(\xi)}{\partial \xi_i \partial \xi_k} \\
\frac{\partial^2 l(\xi)}{\partial \xi_j \partial \xi_k}
\end{bmatrix}.
\] (2.12)

We may consider testing the restriction by testing the null hypothesis \(H_0: \xi_2 = 0\) using the likelihood ratio statistic

\[
\hat{\Lambda} = 2l(\hat{\xi}) - 2l(\hat{\xi}_r),
\] (2.13)

where \(l(\cdot)\) is the logarithm of the maximum likelihood function, and \(\hat{\xi}_r\) and \(\hat{\xi}\) are the corresponding maximum likelihood estimates under the null and alternative hypotheses, respectively. Under \(H_0\) and for large \(n\), the statistic \(\hat{\Lambda}\) approximately follows the \(\chi^2_{k_2}\) distribution with \(k_2\) degrees of freedom.

### 2.2.5 Shrinkage and Positive Shrinkage Estimators

The shrinkage estimator (SE) of \(\xi\) is defined as:

\[
\hat{\xi}_S = \hat{\xi} + (1 - (k_2 - 2)\hat{\Lambda}^{-1})(\hat{\xi} - \hat{\xi}_r), \quad k_2 \geq 3.
\]

This brings us down to a convex combination function with the form \(\hat{\xi}_S = \lambda\hat{\xi} + (1 - \lambda)\hat{\xi}_r\), where \(\lambda \in [0, 1]\). When \(\lambda = 1\), no shrinkage occurs and the estimates are the same as the UE, and if \(\lambda = 0\) the RE is chosen. The drawback for this estimator is that the factor \((1 - (k_2 - 2)\hat{\Lambda}^{-1})\) can be negative. This happens for small values of \(\hat{\Lambda}\). This phenomenon is known as over-shrinkage. This can be relieved by taking its
positive part, which makes it not only a shrinkage estimator but also a thresholding estimator. The positive part shrinkage estimator (PSE) is defined as:

\[ \hat{\xi}_{S+} = \hat{\xi}_r + (1 - (k_2 - 2)\hat{\Lambda}^{-1})^+(\hat{\xi} - \hat{\xi}_r), \quad k_2 \geq 3, \]

and where \( z^+ = \max(0, z) \).

2.3 Asymptotic Results

2.3.1 Asymptotic Distributional Results: Bias

This subsection deals with the asymptotic features of the proposed estimators of \( \xi \) for the RegARIMA model. Firstly, the asymptotic distributional bias (ADB) of the estimators of \( \xi \) will be discussed in detail, and after that, the asymptotic distributional risk (ADR).

Under nonlocal (fixed) alternatives \( H_a : \xi_2 \neq 0 \), \( \hat{\xi}_S \) and \( \hat{\xi}_{S+} \) asymptotically converge to \( \hat{\xi} \), while \( \hat{\xi}_r \) holds unbounded risk. To capture the purposeful comparisons of the proposed estimators in terms of their biases and risks, we consider the sequence of local alternatives:

\[ K(n) : \xi_2 = \frac{h}{\sqrt{n}}, \quad (2.14) \]

Note that under \( H_0 \), \( h = 0 \), it is implied that \( \xi_2 = 0 \), which is a special case of (2.14). Under local \( K(n) \), the following theorem facilitates the derivation and
numerical computation of the ADB and the ADR of the estimators outlined below:

**Theorem 3.1.1.** Under the local alternatives $K(n)$ in (2.14) and the usual regularity conditions Pandher (2022)[50][see, page. 6]

i. $\sqrt{n}\hat{\xi} \xrightarrow{d} N(h, I_{22,1})$ as $n \to \infty$ where $I_{22,1} = I_{22} - I_{21}I_{11}^{-1}I_{12}$ is a positive definite matrix and the information matrix for $\xi = (\xi_1^T, \xi_2^T)^T$

$$I_{\xi_1\xi_2} = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}$$

ii. As $n \to \infty$, the distribution of $\hat{\Lambda}$, converges to a non-central chi-squared distribution $\Psi_{k_2}(x; \Delta)$ with $k_2$ degrees of freedom and the non-centrality parameter,

$$\Delta = h^T I_{22,1} h,$$

and $\Psi_{k_2}(x; \Delta) = P(\chi_{k_2}^2(\Delta) \leq x), x \in \mathbb{R}^+.$$

The outline of a proof is available in Saleh (2006)[58], Chapter 1.

Usually, the shrinkage estimators are biased, however, bias is accompanied by a reduction in variance. We define ADB of an estimator $\xi^*$ as

$$\text{ADB}(\xi^*) = E \left[ \lim_{n \to \infty} \sqrt{n}(\xi^* - \xi) \right].$$

where $\xi^*$ be a generic notation for any of $\hat{\xi}$, $\hat{\xi}_r$, $\hat{\xi}_s$, and $\hat{\xi}_{s+}$.

**Theorem 3.1.2.** Using the above definition of ADB and Theorem 3.1.1, under the local alternatives $K(n)$ in (2.14), and the usual regularity conditions Pandher (2022)[50][see, page. 6], as $n \to \infty$,

$$\text{ADB}(\hat{\xi}_{s+}) = \text{ADB}(\hat{\xi}_s) - I_{11}^{-1}I_{12}h[\Psi_{\nu+4}(\nu, \Delta) - \nu E(Z_1 I(\nu Z_1 > 1))].$$
where $\nu = k_2 - 2$, $Z_1 = \chi_{k_2+2}^{-2}(\Delta)$, $ADB(\hat{\xi}_S) = -\nu I_{11}^{-1} I_{12} h E(Z_1)$, $ADB(\hat{\xi}_r) = -I_{11}^{-1} I_{12} h$, and $\Psi_g(x, \Delta)$ is the distribution function of a non-central chi-square with $g$ degrees of freedom and non-centrality parameter $\Delta$, and

$$E \left( \chi_g^{-2j}(\Delta) \right) = \int_{0}^{\infty} x^{-2j} d\Psi_g(x, \Delta)$$

**Proof.** See the Appendix Pandher(2022)[50].

The constant term $h$ is common to the ADBs of $\hat{\xi}_r$, $\hat{\xi}_S$, and $\hat{\xi}_{S+}$ and the ADBs differ only by a constant factor $\Delta$. It, then, suffices to compare $\Delta$ only. It is clear that the ADB of the $\hat{\xi}_r$ is an unbounded function of $\Delta$. On the other hand, the ADBs of both $\hat{\xi}_S$, and $\hat{\xi}_{S+}$ are bounded in $\Delta$. Since $E(\chi_{k_2+2}^{-2}(\Delta))$ is a decreasing function of $\Delta$, the ADB of $\hat{\xi}_S$ starts from the origin, increases to a maximum, and then decreases towards 0 as $\Delta > 0$. The characteristics of $\hat{\xi}_{S+}$ are similar to $\hat{\xi}_S$.

### 2.3.2 Asymptotic Distributional Results: Risk

We use the following quadratic loss function to obtain the ADRs of the proposed estimators:

$$\mathcal{L}(\xi^*; A) = \left[ \sqrt{n}(\xi^* - \xi) \right]^T A \left[ \sqrt{n}(\xi^* - \xi) \right],$$

where $A$ is a positive semidefinite weight matrix.

In the literature, the researchers considered $A = I$ as an identity matrix in the simulation study. For instance, Gupta (1989) [27] suggested that weight matrix with
different arbitrary diagonal elements can not ensured the outperform of shrinkage estimators to the Full model estimators.

Therefore, the expected loss function is defined as:

\[ E[\mathcal{L}(\xi^*; A)] \equiv ADR(\xi^*, \xi; A) \equiv ADR(\xi^*, \xi) \]

which is called the risk function. Under the sequence of local alternatives, we define the asymptotic distribution function of an estimator \( \xi^* \) as

\[ G(y) = \lim_{n \to \infty} P \left[ \sqrt{n}(\xi^* - \xi) \leq y | K(n) \right], \]

where \( G(y) \) is nondegenerate distribution function for the estimators we consider. We define the asymptotic distributional quadratic risk (ADR) by:

\[
ADR(\xi^*, A) = \int \cdots \int y^\top A y dG(y) = \text{trace}(A \Gamma),
\]

(2.15)

where \( \Gamma = \int \cdots \int y y^\top dG(y) \) is the dispersion matrix for the distribution \( G(y) \) and \( \Gamma \) is the asymptotic covariance matrix of \( \xi^* \).

An estimator \( \xi^* \) is then said to dominate an estimator \( \xi^0 \) asymptotically if \( ADR(\xi^*; \xi) \leq ADR(\xi^0; \xi) \). If, in addition, \( ADR(\xi^*; \xi) < ADR(\xi^0; \xi) \) for at least some \( (\xi, A) \), then \( \xi^* \) strictly dominate \( \xi^0 \).

**Theorem 3.1.3.** Under the local alternatives \( K(n) \) in (2.14) and the usual regularity conditions Pandher (2022)[50][see, page. 6], as \( n \to \infty \), we obtain the ADR functions
of the proposed estimators by virtue of the following theorem:

\[
ADR(\hat{\xi}_{S+}; A) = ADR(\hat{\xi}_S; A) - \operatorname{trace}(AC)E[(1 - \nu Z_1)^2 I(\nu Z_1 > 1)] - \gamma^\top A \gamma E[(1 - \nu Z_2)^2 I(\nu Z_2 > 1)] + 2\gamma^\top A \gamma E[(1 - \nu Z_1)I(\nu Z_1 > 1)],
\]

where

\[
ADR(\hat{\xi}_S; A) = ADR(\hat{\xi}; A) - 2\nu \operatorname{trace}(AC)[2\Delta E(Z_2^2) + E(Z_1)] + \nu(\nu + 4)\gamma^\top A \gamma E(Z_2^2),
\]

\[
Z_2 = \chi_{k_2+4}(\Delta), \quad ADR(\hat{\xi}; A) = ADR(\hat{\xi}; A) - \operatorname{trace}(A\Phi) + \gamma^\top A \gamma,
\]

and

\[
C = I_{11}^{-1} I_{12}^{-1} I_{21} I_{11}^{-1}.
\]

\[
\text{Proof.} \quad \text{See the Appendix Pandher(2022)[50].}
\]

### 2.4 Simulation Results

This section deals with the simulation study to review the comparable performance of \(\hat{\xi}_r, \hat{\xi}_S, \text{ and } \hat{\xi}_{S+}\) with respect to \(\hat{\xi}\). We consider three scenarios of models for generating data. We explain these models as below:

- **Simulated model:** We consider the following simulation model with ARIMA(1, 1, 1)
error for generating responses

\[ y_t^* = x_t^* \beta + w_t, \]  

(2.16)

where \( \omega_t = \phi_1 \omega_{t-1} + a_t - \theta_1 a_{t-1} \) with \( \omega_t = \nabla \eta_t \). Also \( y_t^* = \nabla y_t \), \( x_t^* \top = \nabla x_t \top \) with \( t = 1, 2, \cdots, n \). We take \( \phi_1 = -0.49, \ d = 1, \ \theta_1 = -0.79, \ \sigma = 1 \) and three covariates with regression coefficients \(-0.95, 1.05, 1.9\). Therefore, \( \beta_{\text{sim}} \) is a vector of regression coefficients, and all the model parameters together are \( \xi_{\text{sim}} = (-0.49, -0.79, \beta_{\text{sim}} \top) \top = (-0.49, -0.79, -0.95, 1.05, 1.9) \top \). Finally, the response is generated from simulated model.

**Unrestricted model:** The unrestricted model specified in (2.2) with \( \hat{\xi} = (\xi_1 \top, \xi_2 \top) \top \), \( \xi_1 = \xi_{\text{sim}} \), and \( \xi_2 \) treated as vector of regression coefficients for \( k_2 \) additional covariates. In this setting, the response is generated from the simulation model in which we consider the true \( \xi_2 \) as a zero vector. Therefore, the true values of the parameters are \( \xi_1 = (-0.49, -0.79, -0.95, 1.05, 1.9) \top \) and \( \xi_2 = \beta_2 = (0, 0, 0, 0, ..., 0) \top \). We consider six values: \( k_2 = 4, 8, 12, 15, 18 \) and 21.

**Restricted model:** The restricted model is established by considering a limitation \( \xi_2 = 0 \) in the unrestricted model. the restricted model is not significantly different from the unrestricted model. In this situation, \( \Delta = ||\hat{\xi} - \xi^*||^2 = 0 \), where \( || \cdot || \) is the Euclidian norm. To explore the behavior of the shrinkage
estimators when the restricted model is substantially different from the unrestrict
ed model, we consider \( \xi_2 = (\sqrt{c}, 0, ..., 0)^\top \) so that \( \Delta = ||\hat{\xi} - \hat{\xi}_r||^2 = c \), where \( c \) is a positive constant. Note that \( \Delta \) is the difference between the restricted and the unrestricted model in the spirit of the local alternative (2.14), and the performance of the shrinkage estimators is assessed under both \( H_0 : \Delta = 0 \) and \( H_1 : \Delta = c \), for \( 0 < c < 1.5 \).

All the covariates generated from the standard normal distributions. We use \( n = 500, 600, 700, \) and \( 800 \) to investigate the impact of \( n, k_2, \) and \( \Delta \) on the shrinkage estimators.

Here, we consider \( \Delta = (0, 0.03, 0.07, 0.10, 0.15, 0.3, 0.55, 1, 1.5) \).

<table>
<thead>
<tr>
<th></th>
<th>n = 500</th>
<th>n = 600</th>
<th>n = 700</th>
<th>n = 800</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k_1 = 3, k_2 = 4 )</td>
<td>( k_1 = 3, k_2 = 4 )</td>
<td>( k_1 = 3, k_2 = 4 )</td>
<td>( k_1 = 3, k_2 = 4 )</td>
</tr>
<tr>
<td>RE</td>
<td>1.875</td>
<td>1.847</td>
<td>1.808</td>
<td>1.829</td>
</tr>
<tr>
<td>SE</td>
<td>1.312</td>
<td>1.309</td>
<td>1.293</td>
<td>1.312</td>
</tr>
<tr>
<td>PSE</td>
<td>1.405</td>
<td>1.416</td>
<td>1.417</td>
<td>1.407</td>
</tr>
<tr>
<td></td>
<td>( k_1 = 3, k_2 = 21 )</td>
<td>( k_1 = 3, k_2 = 21 )</td>
<td>( k_1 = 3, k_2 = 21 )</td>
<td>( k_1 = 3, k_2 = 21 )</td>
</tr>
<tr>
<td>RE</td>
<td>5.63</td>
<td>5.945</td>
<td>5.761</td>
<td>5.690</td>
</tr>
<tr>
<td>SE</td>
<td>3.989</td>
<td>4.06</td>
<td>3.856</td>
<td>3.884</td>
</tr>
<tr>
<td>PSE</td>
<td>4.398</td>
<td>4.541</td>
<td>4.386</td>
<td>4.442</td>
</tr>
</tbody>
</table>

Table 2.1: RMSEs of \( \hat{\xi}_r, \hat{\xi}_S, \) and \( \hat{\xi}_{S+} \) with respect to \( \hat{\xi} \) when the restricted parameter space is correct (\( \Delta = 0 \))
Figure 2.1: RMSEs of RE(\(\hat{\xi}_r\)), SE (\(\hat{\xi}_S\)), and PSE (\(\hat{\xi}_{S+}\)) with respect to URE(\(\hat{\xi}\)) when the subspace misspecifies \(\Delta \geq 0\), \(n = 500, 600\), and \(k_1 = 3, k_2 = 8, 12, 15, 18\).
Figure 2.2: RMSEs of $\text{RE}(\hat{\xi}_r)$, $\text{SE}(\hat{\xi}_S)$, and $\text{PSE}(\hat{\xi}_{S+})$ with respect to $\text{URE}(\hat{\xi})$ when the subspace misspecifies $\Delta \geq 0$, $n = 700, 800$, and $k_1 = 3, k_2 = 8, 12, 15, 18$.

The findings demonstrated in Table 2.1 for $\Delta = 0$ and in the Figures 2.1 and 2.2 for $\Delta \geq 0$. We summarize our findings as below:

(i) In this simulation study, we generated 1000 replications of the model (2.16) with different combinations of $n$, $k_1$ and $k_2$. We calculated the MSEs of the estimators to evaluate the relative efficiency (RMSE) of the estimators with respect to URE. Thus, a RMSE value exceeding one means the estimators have lower risk than the URE, and a RMSE of less than one has higher risk. When
\( \Delta = 0 \), the RE outshines all the estimators. However, in the neighbourhood of \( \Delta \), the RMSE of the RE decreases and goes to zero whereas the RMSEs of the SE and PSE converges to one, as \( \Delta \) gets large.

(ii) The performance of the PSE is better than the SE, as \( \Delta \) gets close to zero. When \( \Delta \) crosses 0.3, the RMSEs of the SE and PSE are identical and approach one as \( \Delta \) increases. The estimation accuracy of the SE and PSE is better for \( n = 500 \), and 600 as compared to the sample size \( n = 700 \), 800 because the RMSEs of SE and PSE increase with a decrease in the sample size. For instance, in Table 2.1, when \( k_1 = 3, k_2 = 21 \) and \( n = 600 \), RMSEs of SE and PSE are 4.06 and 4.541, while \( n = 700, k_1 = 3, k = 21 \), the RMSEs of SE and PSE are 3.856 and 4.386, respectively.

(iii) By increasing the sample size, the RMSE of RE converges faster to zero, but the RMSE of SE and PSE remain constant with the increase of sample size and \( \Delta > 0 \).
Chapter 3

Generalized Autoregressive Moving Average Models: An Efficient Estimation Approach

Abstract

In this paper, we propose an efficient estimation approach, the so-called pretest and shrinkage approaches in estimating the regression parameters of the generalized autoregressive moving average (GARMA) model, which are pervasive for modelling binary and count time series data. This model accommodates a set of covariates in addition to the ARMA parameters. We want to estimate regression and ARMA parameters when some of the regression parameters may belong to a subspace. We apply the maximum partial likelihood method to obtain the unrestricted maximum partial likelihood estimator (UMPLE) and also the restricted maximum partial likelihood estimator (RMPLE) for the model with parameter
restriction and then present the improved pretest and shrinkage estimators. We establish the asymptotic distributional biases and risks of the proposed estimators and evaluate their relative performance with respect to the UMPLE. The performance of the proposed estimators is investigated using simulation studies. A real data example is provided to illustrate practical usefulness of the estimators.

**Keywords:** ADB and ADR; binary and count time series; GARMA Model; improved pretest and shrinkage; simulation studies

### 3.1 Introduction

Categorical and count time series data are collected in many fields of applications, such as finance, insurance, biomedical, and public health. The statistical research concentrating on such data evolved considerably over the years. Binary time series that correspond to categorical data with two categories occur in many different contexts (e.g., daily records of the presence or absence of an asthma attack on each patient in addition to the measurements of air pollutants, daily temperature and humidity). For count time series data, the health officials record the number of disease cases over a certain time period or the daily number of asthma admissions in a hospital to understand the disease trajectory. This allows us to measure the impact of covariates on time series and capture a suitably dependence among observations. For data with
normal distribution, the classical method is used to fit a linear regression where the errors follow a general autoregressive moving average (ARMA) structure Shumway (2006)[62]. For binary and count time series data, the generalized ARMA Benjamin et al. (2003)[9] and the generalized linear ARMA models Dunsmuir (2015)[16] extend the regression model with ARMA error to a amenable data-driven model for non-normal time series allowing us to model categorical, count and continuous time series data.

The ARMA model originated from the autoregressive (AR) model and the moving average (MA) model. This model is more sophisticated compared to AR and MA when we capture short-range dependence in the time series data, as it usually requires more parameters than AR and MA alone. Additionally, when we use the ARMA model in the Backward shift operator, then the AR and MA polynomials can sometimes share a unique factor, therefore leading to a simpler model. The ARMA model can play a prominent role in modelling short memory high frequency time series data in the field of medicine, economics, and many other areas of research. Many researchers worked on data-based approaches for fitting either variations or special cases of generalized ARMA (GARMA) models under the structure of Benjamin et al. (2003)[9]. For example, Moller (2020)[41] proposed a generalized discrete ARMA approach for stationary multivariate or univariate time series data of integers, counts, and compositional data. Albarracin (2019)[3] focused on GARMA models where the
conditional mean depends on covariates and also lagged values of the response in the set of regressors to model the serial dependence. They made inference about the effects of covariates on the mean response and also took into account the serial dependence. An observation-driven ARMA model for count time series data with covariate as implemented in the \texttt{glarma} R package Dunsmuir (2015)[16]. This package models the serial dependence using the GARMA class of observation-driven models and provides valid inference for the regression model components. Woodard (2011)[72] presented some general results on stationary and ergodic solutions for some observation-driven GARMA models for binary, count, or other discrete-valued data which are similar to those available in Benjamin et al.[9]. Rocha (2009) [57] proposed beta autoregressive moving average models for rates and proportions over time, a continuous variable in the interval (0, 1). Startz (2008)[63] introduced binomial ARMA models for binary data generated from U.S. recession time series data and results compared to traditional Markov models. Benjamin (2003)[9] presented a class of GARMA models for non-Gaussian time series data and discussed the relation to the other models such as AR, MA, and GARCH models.

In this paper, we consider the problem of improved pretest and shrinkage estimation for GARMA when there are many covariates under investigation. Due to over-modelling, researchers are facing a difficult task to get an optimum choice of the
active covariates from the model and their aim to decrease the number of inactive co-
variates for increasing predictive capability. This goal can be fulfilled by introducing
the improved pretest and shrinkage estimators. To form these estimators, we apply
a two-step approach for fitting GARMA models where many of the regression coeffi-
cients are zero or under linear restriction. The first step provides naive selection of
nonzero regression coefficients based on backward elimination with Akaike informa-
tion criterion. Next, a constraint on the regression parameter vector is to be placed
by using the inactive covariates. In the GARMA models, let the regression parameter
vector be $\gamma$ of dimension $l \times 1$, where $l$ is the number of covariates. We define a
subspace that satisfies a set of $l_2$ linear restrictions $R\gamma = h$, where $R$ is $l_2 \times l$
matrix of rank $l_2$ and $h$ is a $l_2 \times 1$ vector of constants. In particular, if $R = [O, I]$, where $O$
and $I$ are zeros and identity matrices of orders are $l_2 \times l_1$ and $l_2 \times l_2$ respectively, then
we can divide the full parameter vector into sub-vectors $\gamma = (\gamma_1^T, \gamma_2^T)^T$ with order
$l_1 \times 1$ of $\gamma_1$ and $l_2 \times 1$ of $\gamma_2$ respectively, such that $l = l_1 + l_2$. In the second step, we
combine the full model and the model with restriction in an optimal way, in order to
achieve the improved pretest and shrinkage estimators.

Shrinkage estimators have been proposed and studied as efficient solutions in
the past research works related to ARMA and other time series models. Hossain
(2016)\cite{32} introduced the pretest and James-Stein shrinkage estimators for generalized
linear model for the time series data in which the conditional mean of the counts and
binary responses are specified as the product of a deterministic exponential function of known covariates and a latent stationary process. Nkurunziza (2015)[44] proposed James-Stein shrinkage estimators of the memory parameters of a stationary Gaussian times series process when some of the parameters are under the linear restrictions and showed the proposed estimators are more efficient than the maximum likelihood estimators. Pchelintsev (2013)[53] considered the problem of shrinkage estimation for a continuous time regression with a dependent non-Gaussian noise of pulse type, and the noise is specified by the Ornstein–Uhlenbeck process. More discussion of the pretest and shrinkage estimation methodologies and their biases and risks can be found in Hossain (2016)[32], Hossain (2015)[31], Raheem(2012)[55], and An (2009)[6].

The reminder of this article is organized as follows. In Section 3.2, we present the model specification of GARMA. In Section 3.3, we set up an estimation method for fitting GARMA models, the partial score vector, and conditional information matrix. Section 3.4 examines the asymptotic distributional biases and risks of the proposed estimators. In Section 3.5, we introduce a Monte Carlo simulation study to evaluate the numerical performance of the maximum partial likelihood estimator. An application of GARMA models to S&P 500 data is presented in Section 3.6. Section 3.7 closes the article with concluding remarks.
3.2 Model specification of GARMA

In the GARMA models, the distribution of \( \{y_t, t \geq 1\} \) process, conditioned on the past information \( \mathcal{F}_{t-1} = \sigma \{y_{t-1}, \ldots, y_1, x_t, \ldots x_1\} \), belongs to the exponential family, where \( \mathcal{F}_{t-1} \) is \( \sigma \)-field up to time \( t-1 \) which generated through previous values of response variables \( \{y_t\} \) and possibly current and previous values of the covariates \( \{x_t\} \). It is a generalization of autoregressive moving average models to exponential family distributions, providing direct treatment of count and binary-valued data, and so on. It assumes the conditional density function of \( y_t \) is:

\[
f(y_t; \theta_t, \psi|\mathcal{F}_{t-1}) = \exp \left( \frac{y_t \theta_t - b(\theta_t)}{a_t(\psi)} + c(y_t; \psi) \right), \quad t = 1, 2, \ldots, n. \tag{3.1}
\]

where \( \theta_t \) is the natural parameter, \( \psi \) is a dispersion parameter, and \( a_t(.), b(.), \) and \( c(.) \) are specific functions that represents a particular distribution from the exponential family such as Poisson and Binomial distributions. Following McCullagh (1989)[40], \( \mu_t = E(y_t|\mathcal{F}_{t-1}) = b'(\theta_t) \) and \( \sigma_t^2 = \text{Var}(y_t|\mathcal{F}_{t-1}) = a_t(\psi)b''(\theta_t) \) are conditional mean and variance of \( y_t \), respectively.

For instance, if \( y_t \) follows Poisson distribution instead of the exponential family distribution then \( a_t(\psi) = 1, \theta_t = \ln \mu_t, \quad b(\theta_t) = \exp(\theta_t), \quad c(y_t; \psi) = -\ln(y_t!) \) and if \( y_t \) follows binomial distribution, then \( a_t(\psi) = m_t, \theta_t = \ln(\pi_t/(1-\pi_t)), \quad b(\theta_t) = \ln(1+\exp(\theta_t)), \quad c(y_t; \psi) = -\ln\left(\binom{m_t}{y_t}\right) \), where \( m_t \) is the number of trials at time \( t \) and \( \pi_t = p(y_t = 1|\theta_t) \) is the success probability.
As in the classical GLM, the link function $g(.)$ relates the linear predictor $\eta_t$ to the mean $\mu_t$, where $g(.)$ is a real valued monotonic and twice differentiable function.

The linear predictor of the GARMA models can be written as:

$$g(\mu_t) = \eta_t = x_t^T \gamma + \sum_{j=1}^{p} \phi_j H(y_{t-j}, x_{t-j}, \gamma) + \sum_{j=1}^{q} \delta_j D(y_{t-j}, \mu_{t-j}),$$  

(3.2)

where $x_t$ is a $l \times 1$ covariate vector, $\gamma = (\gamma_1, \cdots, \gamma_l)^T$ is a corresponding parameter vector, $H$ and $D$ are functions representing the AR and MA terms and $\phi = (\phi_1, \phi_2, \cdots, \phi_p)^T$ and $\delta = (\delta_1, \delta_2, \cdots, \delta_q)^T$ are the AR and MA parameters, respectively. The MA error term $D$ could be deviance residuals, Pearson residuals, residuals measured on the original scale or residuals on the predictor scale. Practically, the linear predictor (3.2) can be rewritten as:

$$g(\mu_t) = \eta_t = x_t^T \gamma + \sum_{j=1}^{p} \phi_j (g(y_{t-j}) - x_{t-j}^T \gamma) + \sum_{j=1}^{q} \delta_j (g(y_{t-j}) - \eta_{t-j}).$$  

(3.3)

Model (3.1) with (3.3) can be termed as GARMA models of order $(p, q)$ as it has both AR and MA terms (Benjamin et al. (2003)[9], Kedem (2002) [36]).

For certain link functions $g$, it may be required to replace $y_{t-j}$ by $y_{t-j}^*$ in (3.3) to avoid the non-existence of $g(y_{t-j})$ for certain values of $y_{t-j}$. For example, a Poisson conditional distribution for $y_t$ given $\mathcal{F}_{t-1}$ in (3.1) gives the Poisson GARMA($p, q$) model along with $g(\mu_t) = \log(\mu_t)$ in (3.3) can be expressed as:

$$\log(\mu_t) = x_t^T \gamma + \sum_{j=1}^{p} \phi_j (\log(y_{t-j}^*) - x_{t-j}^T \gamma) + \sum_{j=1}^{q} \delta_j (\log(y_{t-j}^*/\mu_{t-j})),$$  

(3.4)
where \( y_{t-j}^* = \max(y_{t-j}, c) \) and \( 0 < c < 1 \). Similarly, if \( y_t \) follows conditional binomial 
\( B(m_t, \mu_t) \) distribution given \( \mathcal{F}_{t-1} \) in (3.1) with \( g(\mu_t) = \logit(\mu_t) = \log(\mu_t/(m_t - \mu_t)) \) in (3.3), then it is called binomial logistic GARMA\((p,q)\) model, where \( y_t^* = \min[\max(y_t, c), (m_t - c)] \). In particular, if \( m_t = 1 \) for \( t = 1, 2, \cdots, n \), it is just a binary logistic GARMA\((p,q)\) model.

### 3.3 Estimation of GARMA models

The estimation of the parameters which index the GARMA models can be conducted through the conditional partial maximum likelihood estimation using iteratively reweighted least squares (IRLS) algorithm. Given \( n \) successive observation \( \{(y_t, x_t^\top), t = 1, 2, \cdots, n\} \) on the response series, the likelihood function is constructed as the product of conditional densities of \( y_t \) given \( \mathcal{F}_{t-1} \). Let us denote the \((l+p+q)\) dimensional vector of parameters as \( \beta = (\gamma^\top, \xi^\top)^\top \) with \( \xi^\top = (\phi^\top, \delta^\top)^\top \). Suppose the covariates \( x_t \) are stochastic, then the estimation of \( \beta \) is done by maximizing the partial likelihood (Fahrmeir (2001)[18, pg. 194]) with IRLS algorithm. The partial likelihood (PL) of the observed series is:

\[
PL(\beta) = \prod_{t=1}^{n} f(y_t; \theta_t, \psi|\mathcal{F}_{t-1}).
\]
Then from (3.1), the log-partial likelihood $\ell(\beta)$, is given by:

$$
\ell(\beta) = \sum_{t=1}^{n} \log f(y_t; \theta_t, \psi | \mathcal{F}_{t-1}) = \sum_{t=1}^{n} \left( \frac{y_t \theta_t - b(\theta_t)}{a_t(\psi)} + c(y_t; \psi) \right).
$$

To give the importance of the dependence of the partial likelihood on $\gamma$, we have used the notation $u(\cdot) \equiv (g \circ \mu(\cdot))^{-1} = \mu^{-1}(g^{-1}(\cdot))$, where the small circle $\circ$ represents the composition of functions $g$ and $\mu$ so that $\theta_t = u(x_t^\top \gamma)$. Since the log-likelihood function (3.5) is the sum of the log of conditional densities $f(y_t; \theta_t, \psi | \mathcal{F}_{t-1})$, then the log-likelihood of the data $\{y_{m+1}, y_{m+2}, \cdots, y_n\}$ conditional on the first $m$ observations $\{y_1, y_2, \cdots, y_m\}$ and on $\eta_t = g(y_t^*)$ for $t = 1, 2, \cdots, i$ where $i = \max\{p, q\}$, and $m \geq i$ can be written as:

$$
\ell(\beta) = \sum_{t=m+1}^{n} l_t, \quad (3.6)
$$

The unrestricted maximum partial likelihood estimator (UMPLE) of $\beta$ is obtained by maximizing the conditional log-likelihood function $\ell(\beta)$ in (3.6) using IRLS algorithm. This procedure is similar to the maximization procedure used for classical generalized linear models.
Define two matrices $J_1$ and $J_2$:

$$J_1 = \begin{bmatrix}
g(y_m) - x_m^\top \beta & \cdots & g(y_{m+1}) - x_{m+1-1}^\top \beta \\
g(y_{m+1}) - x_{m+1}^\top \beta & \cdots & g(y_{m+2}) - x_{m+2-1}^\top \beta \\
\vdots & \vdots & \vdots \\
g(y_{n-1}) - x_{n-1}^\top \beta & \cdots & g(y_n) - x_{n-p}^\top \beta
\end{bmatrix}$$

$$J_2 = \begin{bmatrix}
g(y_m) - \eta_m & \cdots & g(y_{m-q}) - \eta_{m-q} \\
g(y_{m+1}) - \eta_{m+1} & \cdots & g(y_{m+2-q}) - \eta_{m+2-q} \\
\vdots & \vdots & \vdots \\
g(y_{n-1}) - \eta_{n-1} & \cdots & g(y_{n-q}) - \eta_{n-q}
\end{bmatrix}.$$  

The linear predictor in (3.3) for the observations, $y_{m+1}, y_{m+2}, \cdots, y_n$ can be rewritten in matrix form $\eta_t = J^\top \beta$, where $J = [X, J_1, J_2]$, $X$ is the matrix of past covariates, $J_1$ is a matrix of autoregressive terms, and $J_2$ is a matrix of moving average terms.

For the development of conditional score vector, we differentiate the conditional log-likelihood function $\ell(\beta)$ given in (3.6), with respect to $\beta = (\gamma^\top, \phi^\top, \delta^\top)^\top$. More details of the conditional score vector can be found in Bayer(2017)[8] and Kedem (2002)[36, chap. 2]. Now, the partial likelihood estimation of the parameter vector $\beta$ is based on the following partial score function:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{t=m+1}^n \frac{\partial l_t}{\partial \theta_t} \cdot \frac{\partial \theta_t}{\partial \mu_t} \cdot \frac{\partial \mu_t}{\partial \beta_j} + \sum_{t=m+1}^n \frac{y_t - \mu_t(\beta)}{\sigma_t^2(\beta)} \cdot \frac{\partial \mu_t}{\partial \eta_t} \cdot \frac{\partial \eta_t}{\partial \beta_j}, \quad (3.7)$$

and that the calculation of $\frac{\partial \eta_t}{\partial \beta_j}$ is carried out by the following recursive equations for
\( t > \max\{p, q\} \):

\[
\frac{\partial \eta_t}{\partial \gamma_s} = x_{ts} - \sum_{j=1}^{p} \phi_j x_{t-j}s - \sum_{j=1}^{q} \delta_j \frac{\partial \eta_{t-j}}{\partial \gamma_s}, \quad \text{for } s = 1, 2, \cdots, l.
\]

\[
\frac{\partial \eta_t}{\partial \phi_s} = g(y_{t-s}) - x_{t-s}^\top \gamma - \sum_{j=1}^{q} \delta_j \frac{\partial \eta_{t-j}}{\partial \phi_s}, \quad \text{for } s = 1, 2, \cdots, p.
\]

\[
\frac{\partial \eta_t}{\partial \delta_s} = g(y_{t-s}) - \eta_{t-s} - \sum_{j=1}^{q} \delta_j \frac{\partial \eta_{t-j}}{\partial \delta_s}, \quad \text{for } s = 1, 2, \cdots, q.
\]

Finally, let \( A, B \) and \( C \) be the matrices of order \((n-m) \times l, (n-m) \times p, (n-m) \times q\), respectively whose \((t,s)\)th elements are given by \( A_{t,s} = \frac{\partial \eta_{t+m}}{\partial \gamma_s}, B_{t,s} = \frac{\partial \eta_{t+m}}{\partial \phi_s}, \) and \( C_{t,s} = \frac{\partial \eta_{t+m}}{\partial \delta_s} \).

Let \( y = (y_{m+1} - \mu_{m+1}, y_{m+2} - \mu_{m+2}, \cdots, y_n - \mu_n)^\top \) and \( T = \text{diag}(1/\sigma_{m+1}^2(\beta), 1/\sigma_{m+2}^2(\beta), \cdots, 1/\sigma_n^2(\beta)) \). Now we can write the partial score vector as \( U(\beta) = (U_\gamma(\beta)^\top, U_\phi(\beta)^\top, U_\delta(\beta)^\top)^\top \), where \( U_\gamma(\beta) = A^\top Ty, U_\phi(\beta) = B^\top Ty, \) and \( U_\delta(\beta) = C^\top Ty \). We obtained the conditional maximum likelihood estimates by solving the non-linear system \( U(\beta) = 0 \), where \( 0 \) shows the null vector in \( \mathbb{R}^{l+p+q} \). This non-linear system does not have closed form solution. The solution of this non-linear system is obtained using the IRLS process by the method of Fisher scoring algorithm, which is the modification of the Newton-Raphson method. At the \((k+1)\)th iteration, the Fisher scoring method simplifies to:

\[
\hat{\beta}^{k+1} = (J^\top W^k J)^{-1} J^\top W^k z^k, \quad (3.8)
\]

where \( W = \text{diag}\{w_{m+1}, \cdots, w_n\} \) and the dependent variable \( z \) and weights \( W \) are
constructed from \( z = \partial \eta_t / \partial \beta_t^\top + \alpha (y_t - \mu_t) \partial \eta_t / \partial \mu_t \) and \( w_t^{-1} = (\partial \eta_t / \partial \mu_t)^2 v(\mu_t) \), respectively for \( t = m + 1, m + 2, \cdots, n \) and \( 0 < \alpha \leq 1 \). Here \( v(\mu_t) \) is the variance function and \( \alpha \) is the step length that supports the convergence of estimation procedure. The limiting version of (3.8) is called the UMPLE of \( \beta \). The iterative procedure is called the IRLS, since each iteration has the same form as that of weighted least squares but with adjusted weight \( W \) and adjusted dependent variable \( z \).

### 3.3.1 Conditional Information Matrix

The asymptotic distribution of the partial maximum likelihood estimator relies on the Fisher conditional information matrix. To derive this matrix, we need to compute the expected values of all second order derivatives. By differentiating the conditional log-likelihood function given in (3.7), with respect to the \( i \)th element of the parameter vector \( \beta \), we can get for \( i, j \in \{1, 2, \cdots, l + p + q\} \)

\[
\frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j} = \sum_{t=m+1}^{n} \frac{\partial}{\partial \mu_t} \left( \frac{\partial l_t}{\partial \mu_t} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \beta_j} \right) \frac{\partial \mu_t}{\partial \beta_i} \frac{\partial \eta_t}{\partial \beta_i} 
= \sum_{t=m+1}^{n} \left[ \frac{\partial^2 l_t}{\partial \mu_t^2} \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \beta_j} + \frac{\partial l_t}{\partial \mu_t} \frac{\partial}{\partial \mu_t} \left( \frac{\partial \mu_t}{\partial \eta_t} \frac{\partial \eta_t}{\partial \beta_j} \right) \right] \frac{\partial \mu_t}{\partial \beta_i} \frac{\partial \eta_t}{\partial \beta_i}.
\]

Bayer (2017)[8] showed that \( E \left( \frac{\partial l_t}{\partial \mu_t} \bigg| \mathcal{F}_{t-1} \right) = 0 \), thus we can write the above as:

\[
E \left( \frac{\partial^2 l(\beta)}{\partial \beta_i \partial \beta_j} \bigg| \mathcal{F}_{t-1} \right) = \sum_{t=m+1}^{n} E \left( \frac{\partial^2 l_t}{\partial \mu_t^2} \bigg| \mathcal{F}_{t-1} \right) \left( \frac{\partial \mu_t}{\partial \eta_t} \right)^2 \frac{\partial \eta_t}{\partial \beta_i} \frac{\partial \eta_t}{\partial \beta_j} = - \sum_{t=m+1}^{n} \frac{w_t}{\left( g(\mu_t)^2 \right) \partial \beta_i \partial \beta_j}, \tag{3.9}
\]
where \( w_t = \mathbb{E}(\partial^2 l_t/\partial \mu_t^2 | \mathcal{F}_{t-1}) \) and \( g'(\mu_t) = \partial \eta_t/\partial \mu_t \). Let \( W = \text{diag}(w_{m+1}, w_{m+2}, \ldots, w_n) \), then the elements of joint conditional Fisher information matrix for \( \beta \) is:

\[
\mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \gamma_i \partial \gamma_j} \bigg| \mathcal{F}_{t-1} \right) = -A^\top W A; \quad \mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \gamma_i \partial \phi_j} \bigg| \mathcal{F}_{t-1} \right) = -A^\top W B
\]

\[
\mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \gamma_i \partial \delta_j} \bigg| \mathcal{F}_{t-1} \right) = -A^\top W C; \quad \mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \phi_i \partial \phi_j} \bigg| \mathcal{F}_{t-1} \right) = -B^\top W B
\]

\[
\mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \phi_i \partial \delta_j} \bigg| \mathcal{F}_{t-1} \right) = -B^\top W C; \quad \mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \delta_i \partial \delta_j} \bigg| \mathcal{F}_{t-1} \right) = -C^\top W C.
\]

For each component \( \beta_j \) of \( \beta = (\gamma^\top, \phi^\top, \delta^\top)^\top \) and for all \( i, j \in \{1, \ldots, l + p + q\} \) let

\[
G_{\beta_i, \beta_j}(\beta) = \mathbb{E}\left( \frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j} \bigg| \mathcal{F}_{t-1} \right),
\]

for \( i \leq j \) and \( G_{\beta_i, \beta_j}(\beta) = G_{\beta_j, \beta_i}(\beta)^\top \) for \( i > j \). Therefore, the joint conditional information matrix is \( G_n(\beta) \), whose \((i, j)\) element is \( G_{\beta_i, \beta_j}(\beta) \) can be written in the following matrix form:

\[
G_n(\beta) = \begin{bmatrix}
G_{\gamma\gamma} & G_{\gamma\phi} & G_{\gamma\delta} \\
G_{\phi\gamma} & G_{\phi\phi} & G_{\phi\delta} \\
G_{\delta\gamma} & G_{\delta\phi} & G_{\delta\delta}
\end{bmatrix},
\]

where \( G_{\gamma\gamma} = -A^\top W A, G_{\gamma\phi} = -A^\top W B, G_{\gamma\delta} = -A^\top W C, G_{\phi\phi} = -B^\top W B, G_{\phi\delta} = -B^\top W C \) and \( G_{\delta\delta} = -C^\top W C \). We note that the conditional Fisher information matrix is not a block diagonal, hence the parameters are not orthogonal (Cox (1987)[13]).
The following regularity conditions (Kedem (2002)[see 36, chap. 1]) are required to discuss the asymptotic properties of UMPLE, $\hat{\beta}$.

i. The true parameter vector $\beta_0$ is an interior point in the compact parameter space $\Omega \subseteq \mathbb{R}^{l+p+q}$.

ii. The covariate vector $x_{t-1}$ almost surely lies in the compact set $\Gamma \subseteq \Omega$, such that $\Pr(\sum_{t=1}^{n} x_{t-1} x_{t-1}^\top > 0) = 1$. Also $x_{t-1}^\top \gamma$ lies almost surely in the domain of the inverse link function $h = g^{-1}$ for all $x_{t-1} \in \Gamma$ and $\gamma \in \Omega$.

iii. The components of $\partial \ell(\beta)/\partial \beta$ are assumed to be linearly independent.

iv. The log-likelihood function $\ell(\beta)$ is third times differentiable with respect to $\beta$.

v. There exists a probability measure $\varrho$ in $\Omega$ such that $\int_{\Omega} x_{t-1} x_{t-1}^\top \varrho(dx_{t-1})$ is positive definite and such that under (1) and (2) for Borel sets $A \subset \Omega$, $1/n \sum_{t=1}^{n} I[x_{t-1} \in A] \rightarrow \varrho(A)$ in probability as $n \to \infty$, at the true value of $\beta$.

Assumptions 1–4 are usual regularity conditions for deriving asymptotic normality of the conditional UMPLE. Assumption 1 states that the true parameter may belong to the boundary of the parameter space. Assumptions 2-4 guarantee that the third derivative of the log-partial likelihood is continuous with respect to $\beta$. Therefore the conditional information matrix is positive definite and its inverse would exist with probability 1 for large $n$. Assumption 5 states that $G(\beta)$ is a positive definite matrix at the true value of $\beta_0$, and therefore its inverse exists.
**Theorem 1**: Under the regularity conditions of 1-5, the probability that a locally unique conditional maximum likelihood estimator exists with probability converging to one, as \( n \to \infty \). Let \( \hat{\beta} \) of UMPLE based on a given sample and \( \hat{\beta} \) is almost surely convergence to \( \beta \). Then, as \( n \to \infty \)

\[
\sqrt{n}(\hat{\beta} - \beta) \to N_{l+p+q}(0, G^{-1}(\beta)) \quad \text{in distribution.}
\]

**Proof**: Similar proof is given in Bayer (2017)[8] and Moysiadis (2017)[42].

### 3.3.2 The restricted maximum partial likelihood estimator

Consider the following null and alternative hypotheses:

\[
H_0 : R\gamma = h \quad \text{vs.} \quad H_a : R\gamma \neq h, \quad (3.11)
\]

where \( R \) is an appropriate known matrix with full rank \( l_2 \leq l + p + q \) and \( h \) is a \( l_2 \times 1 \) vector of constant. For example, we consider the partition of the full parameter vector \( \beta = (\gamma_1^\top, \gamma_2^\top, \xi^\top)^\top \) with \( \gamma = \gamma_1 + \gamma_2 \), where \( \gamma_1, \gamma_2, \) and \( \xi \) are \( l_1 \times 1, l_2 \times 1, \) and \( (p + q) \times 1 \) vectors, respectively. Note that by letting \( H_0 : \gamma_2 = 0 \), one can test whether \( \gamma_2 = 0 \).

We now form a modified partial log-likelihood function (3.6) under restriction \( R\gamma = h \) in the null hypothesis (3.11) as

\[
\ell(\beta) + \lambda^\top(R\gamma - h),
\]

where \( \lambda \) is a \( l_2 \times 1 \) vector of Lagrange multipliers. This likelihood function can be maximized by using the same methodology discussed in previous Section with an appropriate choice of
matrix $\mathbf{R}$ and the constant vector $\mathbf{h}$. We called this estimator the restricted maximum partial likelihood estimator (RMPLE) and it is denoted by $\hat{\beta} = (\hat{\gamma}^\top, \hat{\xi}^\top)^\top$. Although we obtained the UMLE and RMLE estimates of ARMA parameters $\hat{\beta}$ and $\tilde{\beta}$, respectively, our primary focus is on the regression parameter vector $\gamma$, while the ARMA parameters are treated as nuisance parameters.

3.3.3 Shrinkage and improved pretest estimators

Before defining the estimators, we need to define the test statistic to test the hypothesis that some of the regression parameters are zero or their linear combination are zero. Three types of test statistics are usually used to test the hypothesis (3.11): log-partial likelihood ratio statistic, wald statistic, and the score test. The distributions of three test statistics are asymptotically equivalent for large $n$ and converge to a chi-squared distribution when $H_0$ is true (Kedem (2002)[36]). As the likelihood ratio test is very popular, we use the following log-partial likelihood ratio test to define the improved pretest and shrinkage estimators:

$$\hat{D}_n = 2 \left[ \ell(\hat{\beta}) - \ell(\tilde{\beta}) \right]$$

where $\hat{\beta}$ and $\tilde{\beta}$ be the UMLE and RMLE of $\beta$, respectively. As our primary focus is on regression parameters $\gamma$, next we define the improved pretest and shrinkage estimators for $\gamma$ only.
Firstly, the pretest estimator (PT) of $\gamma$ based on $\hat{\gamma}$ and $\tilde{\gamma}$ is defined as:

$$\hat{\gamma}_P = \hat{\gamma} - I(\hat{D}_n \leq \chi^2_{l_2,\alpha})(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 1,$$

(3.12)

where $I(F)$ is an indicator function of a set $F$, and $\chi^2_{l_2,\alpha}$ is the $\alpha$-level critical value of the chi-square distribution $l_2$ degrees of freedom. Based on the indicator function, $\hat{\gamma}_P$ chooses $\hat{\gamma}$ or $\tilde{\gamma}$ according to whether $H_0$ is accepted or rejected, respectively. It is important to note that $\hat{\gamma}_P$ is bounded and performs better than $\tilde{\gamma}$ in some part of the parameter space. For details, see Judge (1978)[35], Nkurunziza (2013)[43] and Ahmed (2006)[2].

A limitation of the PT estimator is that it is a discontinuous function of $\hat{\gamma}$ and $\tilde{\gamma}$, and is dependent on the choice of $\alpha$ through the critical value $\chi^2_{l_2,\alpha}$. We address this limitation by defining the shrinkage estimator (SE) of $\gamma$ as:

$$\hat{\gamma}_S = \hat{\gamma} + (1 - (l_2 - 2)\hat{D}_n^{-1})(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 3.$$

The SE is the linear combination $\lambda\hat{\gamma} + (1 - \lambda)\tilde{\gamma}$, where $\lambda \in [0, 1]$ is the parameter that determines the extent to which these estimates are pooled together. If $\lambda = 0$, then the RMPLE dominates completely, whereas for $\lambda = 1$, then no shrinkage occurs.

A setback with the SE is that it is not a convex combination of the UMPLE and RMPLE. This has the tendency to cause over-shrinking, which gives the estimator the opposite sign of the UMPLE. To overcome this issue, we use the positive-part
shrinkage estimator (PSE):

\[ \hat{\gamma}_{S+} = \hat{\gamma} + (1 - (l_2 - 2)\hat{D}_n^{-1})^+(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 3, \]

and where \( z^+ = \max(0, z) \).

We can define another pretest estimator by replacing \( \hat{\gamma} \) for \( \hat{\gamma}_S \) in the pretest formula (3.12). We denote this estimator as the improved pretest estimator (IPT) and it is defined by:

\[ \hat{\gamma}_{ipt} = \hat{\gamma} + (1 - (l_2 - 2)\hat{D}_n^{-1})(\hat{\gamma} - \tilde{\gamma})I(\hat{D}_n > \chi_{l_2,\alpha}^2), \quad l_2 \geq 3 \]

\[ = \hat{\gamma}_S - (1 - (l_2 - 2)\hat{D}_n^{-1})(\hat{\gamma} - \tilde{\gamma})I(\hat{D}_n < \chi_{l_2,\alpha}^2). \]

The improved pretest estimator dominates the pretest estimator over a range of parameter values even though we have the limitation \( l_2 \geq 3 \). If \( \chi_{l_2,\alpha}^2 \leq (l_2 - 2) \), \( \hat{\gamma}_{ipt} \) behaves like \( \hat{\gamma}_{S+} \). There are some points in the parameter space where the risk of \( \hat{\gamma}_{ipt} \) crosses the risk of \( \hat{\gamma} \) and hence embraces the similar kind of criticism as being absorbed by \( \hat{\gamma}_P \) (Ahmed (1999)[1]).

### 3.4 Asymptotic results

In this section, we consider the asymptotic properties of the proposed estimators for GARMA models where the subspace \( H_a : R\gamma \neq 0 + d \) may be incorrect, and the ARMA parameters are considered as a nuisance. We derive the asymptotic
distributional bias (ADB) and asymptotic distributional risk (ADR) of the UMPLE, RMLE, SE, PSE, and IPT using the local asymptotic normality Van (1998)[67]. The test statistic $\hat{D}_n$ converges to $\infty$ as $n \to \infty$ for the fixed alternative $H_a : R\gamma \neq 0 + d$, where $d = 0$ and $d = (d_1, d_2, \ldots, d_l) \in \mathbb{R}^l$ is a real fixed vector and the IPT, SE, and PSE have asymptotically same performance and they converge to UMPLE in probability. To evaluate the performance among the estimators, we consider the parameter sequences of the form:

$$K_{(n)} : R\gamma = \frac{d}{\sqrt{n}}. \quad (3.13)$$

The quantity $d/\sqrt{n}$ in (3.13) represents the distance between the local alternative and the subspace $R\gamma = 0$. Under $K_{(n)}$, the following theorem facilitates the derivation and numerical computation of the ADB and the ADR of the estimators outlined below:

**Theorem 2:** Under the local alternatives (3.13) and the regularity conditions in Subsection 3.1, the joint distribution is:

$$\begin{bmatrix} \sqrt{n}(\hat{\gamma} - \gamma) \\ \sqrt{n}(\tilde{\gamma} - \gamma) \\ \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \end{bmatrix} \sim N_{3l} \left( \begin{bmatrix} 0 \\
\vartheta \\
-\vartheta \end{bmatrix}, \begin{bmatrix} \mathcal{G}_{11} & \mathcal{G}_{12} & \mathcal{G}_{13} \\
\mathcal{G}_{21} & \mathcal{G}_{22} & \mathcal{G}_{23} \\
\mathcal{G}_{31} & \mathcal{G}_{32} & \mathcal{G}_{33} \end{bmatrix} \right) ,$$

where $\vartheta = -\Sigma R^T (R\Sigma R^T)^{-1} d$ with $\Sigma = G_{\gamma\gamma}^{-1}$ which is the covariance matrix of $\hat{\gamma}$. Also $\mathcal{G}_{11} = \Sigma, \mathcal{G}_{12} = \mathcal{G}_{21}^T = \Sigma - \Sigma (R\Sigma R^T)^{-1} R\Sigma, \mathcal{G}_{13} = \mathcal{G}_{31}^T = \Sigma R^T (R\Sigma R^T)^{-1} R\Sigma, \mathcal{G}_{22} = \mathcal{G}_{12}, \mathcal{G}_{23} = \mathcal{G}_{32}^T = 0$ and $\mathcal{G}_{33} = \mathcal{G}_{13}$. 
First, we start with the definition of ADB of the aforementioned estimators. Consider a sequence of parameter values $\gamma$ and a sequence of estimators $\hat{\gamma}^*$. As $n \to \infty$, let $\sqrt{n}(\hat{\gamma}^* - \gamma)$ converge to a distribution of some random variable $Z$ with distribution $\tilde{F}_z$. Then the ADB of $\hat{\gamma}^*$ is defined by:

$$
ADB(\gamma^*) = \lim_{n \to \infty} E\left[\sqrt{n}(\gamma^* - \gamma)\right] = \int zd\tilde{F}_z(z).
$$

The expressions of ADB and ADR of the estimators are quite lengthy. To write these expressions in a concise way, let $Z_1 = \chi^2_{l_2+2}(\Delta)$ and $Z_2 = \chi^2_{l_2+4}(\Delta)$ be chi-square random variables with non-central parameter $\Delta = d^\top(R\Sigma R^\top)^{-1}d$. Also, let $\Psi_\nu(x, \Delta) = \Pr(\chi^2_\nu(\Delta) \leq x)$ be the distribution function of a non-central chi-square random variable with $\nu$ degrees of freedom, and let the $\alpha$-level critical value for the central $\chi^2$ distribution with $l_2$ degrees of freedom is $\chi^2_{l_2, \alpha}$. In view of Theorem 2, the asymptotic biases of the estimators give us the following theorem.

**Theorem 3:** Assume that the regularity conditions in Subsection 3.1 are satisfied and $\vartheta = -\Sigma R^\top(R\Sigma R^\top)^{-1}d$. Under the local alternatives $K(n)$ in (3.13), the ADBs of the proposed estimators can be written as $n \to \infty$

$$
\begin{align*}
ADB(\hat{\gamma}) &= \vartheta \\
ADB(\hat{\gamma}_{l^2}) &= \vartheta \left(1 - \Psi_{l_2+2}(\chi^2_{l_2, \alpha}, \Delta) - (l_2 - 2)E\left(Z_1^{-1}I(Z_1 \geq \chi^2_{l_2, \alpha})\right)\right) \\
ADB(\hat{\gamma}_S) &= (l_2 - 2)\vartheta E(Z_1^{-1}) \\
ADB(\hat{\gamma}_{S+}) &= ADB(\hat{\gamma}_S) + \vartheta \Psi_{l_2+2}(l_2 - 2, \Delta) - (l_2 - 2)\vartheta E\left(Z_1^{-1}I(Z_1 < l_2 - 2)\right),
\end{align*}
$$
Proof. See the Appendix.

Remark 1: If \( d = 0 \), then the ABDs of four estimators are null vectors, and there is nothing to compare among ADBs of estimators. For meaningful comparison, let \( d \neq 0 \) and \( \tau = \theta / \Delta \). The ADB of any estimator \( \hat{\gamma}_s \) of \( \gamma \) in the above expressions is a scalar function of \( \Delta \) together with \( \tau \). Now it is reasonable to enlighten the comparison of ADBs of the estimators in terms of scalar multipliers only. The ADB of \( \hat{\gamma} \) is an unbounded function of \( \Delta \) but the ADBs of \( \hat{\gamma}_S \), \( \hat{\gamma}_{ipt} \), and \( \hat{\gamma}_{S+} \) are bounded in \( \Delta \) as \( E(Z_i^{-1}) \) is a decreasing function of \( \Delta \). The ADB of \( \hat{\gamma}_S \) converges to \( \hat{\gamma}_{S+} \) as \( \Delta \to \infty \), but the quadratic bias curve of \( \hat{\gamma}^{S+} \) stays underneath the curve of \( \hat{\gamma}^S \), when \( \Delta \) is on and near zero.

The ADB expressions of all estimators are not in scalar form. To make them scalar, we need to convert them in quadratic form, as this form will be used to calculate the quadratic bias numerically in our simulation study. We now define the quadratic bias (QB) of any estimator \( \hat{\gamma}_s \) as :

\[
QB(\hat{\gamma}_s) = [\text{ADB}(\hat{\gamma}_s)]^\top \Sigma [\text{ADB}(\hat{\gamma}_s)].
\]

Second we define the ADR of the estimators. For this it is necessary to define a quadratic loss function :

\[
L(\hat{\gamma}_s; W) = n ((\hat{\gamma}_s - \gamma))^\top W (\hat{\gamma}_s - \gamma),
\]

where \( W \) is a non-negative positive definite matrix. We may choose this matrix as an identity matrix but other choices of this matrix are also available, for example,
$W = \Sigma$ or a general $W$, which gives a loss function that weight each $\gamma$ differently. An identity matrix will be used in the simulation study. The mean squared error (MSE) matrix for any estimator $\hat{\gamma}_s$ under the quadratic loss function is:

$$\text{MSE}(\hat{\gamma}_s) = \lim_{n \to \infty} \mathbb{E}\left\{ (\sqrt{n}(\hat{\gamma}_s - \gamma))(\sqrt{n}(\hat{\gamma}_s - \gamma))^\top \right\} = \int \int \int \mathbb{E}(\tilde{F}_z(z)) = (3.14)$$

Then the ADR is defined as:

$$\text{AR}(\hat{\gamma}_s; W) = \int z^\top W z d\tilde{F}_z(z) = \text{tr}(W \text{MSE}(\hat{\gamma})). \quad (3.15)$$

**Theorem 4:** Assume that the regularity conditions in Subsection 3.1 are satisfied.
Under the local alternatives $K_{(n)}$ in (3.13), the ADR of the positive shrinkage estimator is:

$$\text{ADR}(\hat{\gamma}; W) = \text{trace}(W\Sigma)$$

$$\text{ADR}(\tilde{\gamma}; W) = \text{ADR}(\hat{\gamma}; W) - \text{trace}(W\mathcal{G}_{12}) + \vartheta^T W \vartheta$$

$$\text{ADR}(\hat{\gamma}_S; W) = \text{ADR}(\hat{\gamma}; W) - (l_2 - 2)\text{trace}(W\mathcal{G}_{13})\left((l_2 - 2)E(Z^{-2}_1) - 2E(Z^{-1}_1)\right)$$

$$+ (l_2 - 2)\left((l_2 - 2)E(Z^{-2}_2) - 2E(Z^{-1}_2 - Z^{-1}_1)\right) \vartheta^T W \vartheta$$

$$\text{ADR}(\hat{\gamma}_{ipt}; W) = \text{ADR}(\hat{\gamma}_S; W) - \text{trace}(W\mathcal{G}_{13}) E\left((1 - (l_2 - 2)Z^{-1}_1)^2 I(Z_1 \leq \chi^2_{l_2,\alpha})\right)$$

$$- \vartheta^T W \vartheta E\left((1 - (l_2 - 2)Z^{-1}_2)^2 I(Z_2 \leq \chi^2_{l_2,\alpha})\right)$$

$$- 2\vartheta^T W \vartheta E\left(Z^{-1}_1 I(Z_1 \leq \chi^2_{l_2,\alpha})\right)$$

$$\text{ADR}(\hat{\gamma}_{S+}; W) = \text{ADR}(\hat{\gamma}_S; W) - \text{trace}(W\mathcal{G}_{13}) E\left((1 - (l_2 - 2)Z^{-1}_1)^2 I(Z_1 < l_2 - 2)\right)$$

$$- E\left((1 - (l_2 - 2)Z^{-1}_2)^2 I(Z_2 < l_2 - 2)\right) \vartheta^T W \vartheta$$

$$+ 2E\left((1 - (l_2 - 2)Z^{-1}_1) I(Z_1 < l_2 - 2)\right) \vartheta^T W \vartheta,$$

where $\vartheta = -\Sigma R^T (R\Sigma R^T)^{-1}d$ and $\mathcal{G}_{13} = \Sigma R^T (R\Sigma R^T)^{-1}R\Sigma$.

**Proof:** See Appendix.

**Remark 2:** If $d = 0$, i.e. $\vartheta = 0$, then the ADRs of $\hat{\gamma}$ and $\tilde{\gamma}$ are $\text{trace}(W\Sigma)$ and $\text{ADR}(\hat{\gamma}; W) - \text{trace}(W\mathcal{G}_{13})$, respectively. It shows that the ADR of $\hat{\gamma}$ is fixed but the ADR of $\tilde{\gamma}$ diverges to infinity, as $d$ in $\vartheta$ moves away from a zero vector. The estimators $\hat{\gamma}_S$, $\hat{\gamma}_{ipt}$, and $\hat{\gamma}_{S+}$ have lower MSE than $\hat{\gamma}$ when $\Delta = 0$. The ADRs of the $\hat{\gamma}_S$, $\hat{\gamma}_{ipt}$, and $\hat{\gamma}_{S+}$ increase as $\Delta > 0$, although they are still more efficient estimators.
than $\hat{\gamma}$. The estimators $\hat{\gamma}_S$ and $\hat{\gamma}_{S+}$ are superior or equal to $\hat{\gamma}$ in the entire parameter space irrespective the value of $\Delta$. In the next section, we conduct a simulation study to compare the proposed estimators with respect to UMPLE.

3.5 Simulation

In order to measure the relative performance of the proposed estimators for the parameters of the GARMA models with respect to the UMPLE, we carried out extensive simulation studies for varying sample sizes and numbers of covariates. We used the following time series model to generate the binary and count time series data:

$$g(\mu_t) = \eta_t = x_t^\top \gamma + \nu_t. \quad (3.16)$$

where $\nu_t = \sum_{j=1}^{p} \phi_j(g(y_{t-j}) - x_{t-j}^\top \gamma) + \sum_{j=1}^{q} \delta_j(g(y_{t-j}) - \eta_{t-j})$ is the GARMA $(p, q)$ process and $g(\cdot)$ is the link functions. If $g$ has the logit link then $\mu_t = \exp(\eta_t)/(1 + \exp(\eta_t))$ and log link then $\mu_t = \exp(\eta_t)$, where $\eta_t$ is defined in (3.16). We investigate the performance of the estimators for sample size $n = 200$, and 300 for binary and count responses. The results are based on 1000 simulations for both responses.

For the binary and count time series models, the response $y_t$ and covariates $x_t$ of a GARMA model (3.16) are generated using the following vector autoregressive moving average model of orders $p = 1$ and $q = 1$ (VARMA $(1, 1)$)

$$x_t = A_1 x_{t-1} + u_t + M_1 u_{t-1}, \quad (3.17)$$
where $A_1$ is a fixed $l \times l$ coefficient matrix of AR coefficients with all eigenvalues of $A_1$ be less than one in absolute value, $M_1$ is the $l \times l$ coefficient matrix of MA coefficients with eigenvalues less than or equal to one in absolute value, and $u_t$ is a zero-mean white noise with a non-singular covariance matrix $\Sigma_u$. Before generating the co-variates $x_t$, we need to supply the $l \times l$ matrices $A_1$ and $M_1$ in the model (3.17), where $l = 8, 11, 14, 18$. We then use the `varima.sim()` function under `portes` R-package (Mahdi (2012)[39]) to simulate non-seasonal data $x_t$ from VARMA $(1, 1)$ process. The `InvertQ()` of this package is used to check the validity of the stationary and invertibility assumptions in the VARMA$(1, 1)$ process.
<table>
<thead>
<tr>
<th></th>
<th>Binary</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 200$</td>
<td>$n = 300$</td>
</tr>
<tr>
<td>$l_1 = 3, l_2 = 8$</td>
<td>2.375</td>
<td>2.215</td>
</tr>
<tr>
<td></td>
<td>2.373</td>
<td>2.054</td>
</tr>
<tr>
<td></td>
<td>1.744</td>
<td>1.712</td>
</tr>
<tr>
<td>SE</td>
<td>1.797</td>
<td>1.615</td>
</tr>
<tr>
<td></td>
<td>1.860</td>
<td>1.808</td>
</tr>
<tr>
<td>PSE</td>
<td>1.953</td>
<td>1.753</td>
</tr>
<tr>
<td></td>
<td>2.136</td>
<td>2.071</td>
</tr>
<tr>
<td>IPT</td>
<td>2.258</td>
<td>1.955</td>
</tr>
<tr>
<td></td>
<td>$l_1 = 3, l_2 = 11$</td>
<td>2.836</td>
</tr>
<tr>
<td></td>
<td>2.950</td>
<td>3.154</td>
</tr>
<tr>
<td>RMPLE</td>
<td>2.110</td>
<td>2.100</td>
</tr>
<tr>
<td></td>
<td>2.134</td>
<td>2.287</td>
</tr>
<tr>
<td>SE</td>
<td>2.215</td>
<td>2.232</td>
</tr>
<tr>
<td></td>
<td>2.404</td>
<td>2.544</td>
</tr>
<tr>
<td>PSE</td>
<td>2.560</td>
<td>2.565</td>
</tr>
<tr>
<td></td>
<td>2.718</td>
<td>2.944</td>
</tr>
<tr>
<td>IPT</td>
<td>2.560</td>
<td>2.565</td>
</tr>
<tr>
<td></td>
<td>$l_1 = 5, l_2 = 6$</td>
<td>1.769</td>
</tr>
<tr>
<td></td>
<td>1.735</td>
<td>1.570</td>
</tr>
<tr>
<td>RMPLE</td>
<td>1.404</td>
<td>1.340</td>
</tr>
<tr>
<td></td>
<td>1.400</td>
<td>1.308</td>
</tr>
<tr>
<td>SE</td>
<td>1.453</td>
<td>1.397</td>
</tr>
<tr>
<td></td>
<td>1.495</td>
<td>1.379</td>
</tr>
<tr>
<td>PSE</td>
<td>1.655</td>
<td>1.562</td>
</tr>
<tr>
<td></td>
<td>1.651</td>
<td>1.489</td>
</tr>
<tr>
<td>IPT</td>
<td>2.201</td>
<td>2.067</td>
</tr>
<tr>
<td></td>
<td>2.070</td>
<td>2.046</td>
</tr>
<tr>
<td>RMPLE</td>
<td>1.704</td>
<td>1.677</td>
</tr>
<tr>
<td></td>
<td>1.681</td>
<td>1.660</td>
</tr>
<tr>
<td>SE</td>
<td>1.777</td>
<td>1.754</td>
</tr>
<tr>
<td></td>
<td>1.798</td>
<td>1.793</td>
</tr>
<tr>
<td>PSE</td>
<td>2.031</td>
<td>1.948</td>
</tr>
<tr>
<td></td>
<td>1.972</td>
<td>1.958</td>
</tr>
</tbody>
</table>

Table 3.1: RMSEs of RMPLE, SE, PSE, IPT for binary and count responses.
Now, we consider a special case of the hypothesis $H_0 : R\gamma = 0$. We partition the coefficients vector into two new sub-vectors $\gamma = (\gamma_1^T, \gamma_2^T)^T$, where $\gamma_1$ and $\gamma_2$ are assumed to have dimensions $l_1 \times 1$ and $l_2 \times 1$, respectively, such that $p = l_1 + l_2$. We are interested in estimating the sub-vector $\gamma_1$ by incorporating the auxiliary information of $\gamma_2$ into the estimation procedure, where the null hypothesis $H_0 : \gamma_2 = \mathbf{0}$ and the dimension of $\mathbf{0}$ varies from 5 to 15. We specify $l_1 = 3, 5$ throughout the simulation study, and set $\gamma_1 = (1.49, -0.75, 1.20)$ and $(1.49, -0.75, 1.20, 0.20, -0.22)$, and a significance level of $\alpha = 0.05$. 
Figure 3.1: RMSEs of RMPLE, SE, PSE, and IPT for binary and Count response models with respect to UMPLE when the subspace misspecifies $\Delta \geq 0$, $n = 200, 300$, and $l_1 = 3, l_2 = 5, 15$.

To investigate the behavior of the proposed estimators, we define a distance between the full model and the restricted model as $\Delta = \| \gamma - \gamma^{(0)} \|^2$, where $\| . \|$ is the Euclidean norm, $\Delta$ is a non-negative number, and $\gamma^{(0)} = (\gamma_1^T, 0^T)^T$. We assume $\Delta = 0, 0.25, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75, 1, 1.50, 2$. Given all the true parameter values, the covariates $x_t$, and $\Delta$, we generate the binary and count responses.
To compare the proposed estimators with respect to the UMPLE, we have calculated the mean squared error (MSE) for all estimators by simulation using 1000 replications. To simplify the presentation, we normalize the MSE of the proposed estimators by that of the UMPLE, which we call relative MSE (RMSE). Thus, a RMSE value exceeding one means the proposed estimators have lower risk than the UMPLE, and a RMSE less than one has higher risk. For the purpose of better visualization, we summarize the results in the following Table 3.1 and Figures 3.1 - 3.2.
3.5.1 ADR analysis when $\Delta = 0$

We have provided the simulated MSE for the binary and count time series models in Table 3.1 for $n = 200, 300$. This table shows that the proposed estimators improve, in terms of RMSE on the UMLE. This improvement is substantial for the number of covariates 14 and 16 in the model when the insignificant number of covariates varies...
from 6 to 11. For $\Delta = 0$, the RMPLE performs the best because of its unbiasedness. The IPT is superior to the SE and PSE irrespective of the number of inactive covariates. If the sample size increases, the RMSE of the proposed estimator decreases. The SE and PSE are outperformed by the UMPLE in terms of risk, but the PSE performs better than the SE. The performance of the estimators for the count response model is better than the performance of estimators for the binary response model, except for some sampling fluctuations in the simulated data.
Figure 3.3: MSEs of URMLE, RMPLE, SE, PSE, and IPT for binary and count response models when the subspace misspecifies $\Delta \geq 0$, $n = 200, 300$, and $l_1 = 3, l_2 = 5, 15$.

3.5.2 ADR and MSE analyses when $\Delta \geq 0$

The main reason for this subsection is to evaluate the proposed estimators for $\Delta > 0$, that is, when $H_0 : \gamma_2 \neq 0$. For $0 \leq \Delta \leq 2$ and $n = 200, 300$, Figures 3.1-3.2 show that the RMPLE outperforms the other estimators; however, as the sample size increases to $n = 300$, the RMPLE outperforms the other estimators up until $\Delta < 0.4$ for binary and count response models. As $\Delta$ increases, the RMSE of
RMPLE becomes closer to zero. It also shows that the RMSE of RMPLE for count response model converges to zero faster than the RMSE of RMPLE for binary response model. This suggests that severe departures from the null hypothesis $H_0$ imposed on the both models lead to inefficient estimates of parameter vector $\gamma$. However, this is not the situation for the other estimators. The RMSE for the IPT, SE, and PSE stays equal to or above one and approaches the reference line RMSE=$1$ except some sampling fluctuations. That is, Figures 3.1-3.2 show convincingly that the SE and PSE significantly dominates the UMPLE. This is in agreement with the asymptotic results in Section 4.
Figure 3.4: MSEs of URMLE, RMPLE, SE, PSE, and IPT for binary and count response models when the subspace misspecifies $\Delta \geq 0$, $n = 200, 300$, and $l_1 = 5, l_2 = 3, 13$

To explore the effect of sample size and insignificant covariates, we make a comparison between the MSE curves (Figures 3.3-3.4). For fixed $l_2$, it shows that the larger the sample size, the lower the MSE of the estimates because of decreasing the bias and variance. These figures also show that the MSE is lower for the count response model than the binary response model that is, more fluctuation in the estimates of coefficients of binary response model. We also see that as $l_2$ increases with $n$ fixed,
the MSE of the estimates is higher. The IPT estimate has lower MSE compared to the SE and PSE. The most sensitive estimator is RMPLE, which has quite low MSE for very small $\Delta$ values but a very large MSE for moderate and large $\Delta$ values. On the other hand, the SE and PSE have approximately lower MSE compared to the MSE of UMPLE.

### 3.6 Real data example: binary response

In this section, we demonstrate the improved pretest and shrinkage estimation strategies using the S&P500 stock market data with a binary response. Empirical evidence suggests that the forecasting predictability of excess stock market returns based on macroeconomic variables ranges from weak to non-predictable. Christoffersen (2006)[11] pointed out that the sign predictability may exist even though the mean predictability is close to zero. This sign predictability is very important for creating profitable investment strategies. Nyberg (2011) [48] compared dynamic binary probit models with alternative models in terms of predictive performances, such as exogenous ARMA models and models where forecasts of the asset return volatility are employed to produce sign forecasts for excess stock returns and found that his model outperformed a number of other models such as exogenous ARMA models.

We fitted a binary GARMA model to analyze the monthly closing price of the S&P500 stock market index and the 3-month U.S. Treasury Bill rate ($B_t$) covering
the period from December 1994 to February 2021. The monthly closing price of the S&P500 stock and the 3-month U.S. Treasury Bill rate can be found from the link https://fred.stlouisfed.org/series/TB3MS. The binary response is defined by \( y_t = 1 \) if \( r_t > 0 \), and \( y_t = 0 \) otherwise, where:

\[
r_t = 100 \times \log(\frac{P_{C_t}}{P_{C_{t-1}}}) - B_t, \quad t = 1, 2, \ldots, 315
\]
is the binary stock return indicator generated from the value of S&P500 stock index \((P_{C_t})\) and three-month US Treasury Bill rate \((B_t)\).

Hossain (2016)[32] modeled this data (covers the period from January 1968 to December 2006) using time series models following GLMs, and they used \( \eta_t = x_{t-1}^\top \gamma, \quad t = 1, \cdots, n \) with five lagged responses of covariates, and the covariate process \( x_{t-1} \) was taken to follow a Vector Autoregressive model of order 1. Here we consider the covariate process \( x_{t-1} \) as ARMA model with logit link. Let \( x_{t-1} = (1, P_{C_{t-1}}, P_{C_{t-2}}, \cdots, P_{C_{t-7}}, B_{t-1}, B_{t-2}, \cdots, B_{t-7})^\top \), is the vector of covariates using seven lagged responses, and regression parameters vector \( \gamma = (\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_{14})^\top \).

Therefore, the total number of observations are \( n = 308 \) with \( t = 8, 9, \cdots, 315 \).

Figure 3.5 shows a time series plot of \( r_t, P_t, \) and \( TB_t \). The extremely low return shows at index 1 in Figure 3.5(a) on December, 1994. Observe that the series \( r_t \) is often negative; the proportion of zeroes is 77%. Figures 3.5(b) and 3.5(c) are plots of the raw data, which exhibit a positive and negative trend, respectively.
Initially, the GLM model is applied to find the orders of AR and MA processes for the GARMA model. The ACF and partial ACF plots from fitted GLM indicate that the coefficient of AR at lag 1 is significant and the coefficients of MA at lag 7 and lag 9 are significant. Thus, a GARMA model is fitted after considering the autocorrelations. We find that the Pearson residual autocorrelations are all non-significant as presented in Figure 3.6. We have then applied the backward estimation procedure to find a restricted GARMA model and found that $PC_{t-1}$, $PC_{t-5}$, $PC_{t-6}$, and $PC_{t-7}$ are significant for the likelihood of access stock return. Figure 3.6 shows the ACF and partial ACF plots of the Pearson residuals of the GARMA model. These
plots suggest that the residuals behave as white noise. We then formed a restriction on
the insignificant covariates for improved pretest and shrinkage estimators by putting
$\gamma_2 = (\gamma_2, \gamma_3, \gamma_4, \gamma_8, \ldots \gamma_{14})^T = (0, 0, 0, 0, 0, 0, 0, 0, 0)^T$, $k = 14$, $l_1 = 4$, and
$l_2 = 10$.

![ACF and partial ACF for Binary response.](image)

Figure 3.6: ACF and partial ACF for Binary response.

We apply bootstrapping technique to compute point estimates, standard errors,
and the RMSEs of the proposed estimators. Only significant coefficients of the
GARMA model are reported in Table 3.2. We apply the following bootstrap al-
gorithm:

i. Set the bootstrap sample of size $n = 308$.

ii. Fit the GARMA model and obtain the UMLE $\hat{\gamma}_1$.

iii. Calculate the fitted values of the model, $\hat{\mu}_t$. 
iv. Generate bootstrap replicates \( \{y_t^*\}_{t=1}^n \) with mean \( \hat{\mu}_t \).

v. Estimate \( \gamma_1 \) based on \( \{y_t^*\}_{t=1}^n \) samples.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>( \hat{\gamma}_1 )</th>
<th>( \hat{\gamma}_5 )</th>
<th>( \hat{\gamma}_6 )</th>
<th>( \hat{\gamma}_7 )</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMPLE</td>
<td>-0.050</td>
<td>0.570</td>
<td>-0.670</td>
<td>0.390</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.420)</td>
<td>(0.420)</td>
<td>(0.300)</td>
<td></td>
</tr>
<tr>
<td>RMPLE</td>
<td>-0.013</td>
<td>0.250</td>
<td>-0.510</td>
<td>0.300</td>
<td>6.570</td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.270)</td>
<td>(0.330)</td>
<td>(0.25)</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>-0.010</td>
<td>0.460</td>
<td>-0.62</td>
<td>0.036</td>
<td>1.965</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.360)</td>
<td>(0.390)</td>
<td>(0.280)</td>
<td></td>
</tr>
<tr>
<td>PSE</td>
<td>-0.020</td>
<td>0.430</td>
<td>-0.600</td>
<td>0.350</td>
<td>1.968</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.350)</td>
<td>(0.390)</td>
<td>(0.280)</td>
<td></td>
</tr>
<tr>
<td>IPT</td>
<td>-0.020</td>
<td>0.430</td>
<td>-0.600</td>
<td>0.350</td>
<td>1.960</td>
</tr>
<tr>
<td></td>
<td>(0.028)</td>
<td>(0.360)</td>
<td>(0.380)</td>
<td>(0.270)</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.2: Estimates (first row \( \times 10^{-2} \)) and standard errors (second row \( \times 10^{-2} \)) for significant covariates \( PC_{t-1}(\hat{\gamma}_1) \), \( PC_{t-5}(\hat{\gamma}_5) \), \( PC_{t-6}(\hat{\gamma}_6) \), and \( PC_{t-7}(\hat{\gamma}_7) \)

The RMSEs of RMPLE, SE, PSE, and IPT with respect to UMPLE are 6.57, 1.965, 1.968, and 1.96, respectively which corroborate our analytical and numerical findings for a binary case.
3.7 Concluding remarks

The aim of this paper is to develop the improved pretest and shrinkage estimation approaches for estimating the coefficients of covariates in GARMA models when some coefficients may be under linear restrictions. We have used the GLM framework to fit UMPLE and RMPLE of GARMA models, when the responses are binary and count. The partial maximum likelihood algorithm described is within the IRLS framework to fit these models. We have then optimally combined the UMPLE and RMPLE to get the shrinkage estimators. We have discussed the asymptotic properties of the improved pretest and shrinkage estimators. Our theoretical findings demonstrate that the improved pretest and shrinkage estimators outperform the UMPLE and perform relatively better than the RMPLE in a wide range of the parameter space.

Our findings indicate that RMPLE outshines the SE, PSE, and IPT at and near the null hypothesis (i.e., $\Delta = 0$), but the performance diminishes if the $\Delta$ value increases. However, for a large number of sparse predictors $l_2$, while keeping $l_1$ and $n$ fixed, SE and PSE are less efficient than IPT. It is shown that the SE and PSE are robust and uniformly superior, in terms of mean squared error, relative to the UMPLE subject to sampling fluctuations. But PSE is uniformly better than the SE. The performance of SE, PSE and IPT for the count GARMA model is better than the binary GARMA model, except for some sampling fluctuations in the simulated data.
Through our real data analysis, we have demonstrated the improved pretest and shrinkage estimation approaches and it shows that the results strongly corroborate the findings of the simulation study and suggest the use of the shrinkage estimator when no prior information about the parameter subspace is available. We anticipate that these estimation methods can extend other types of models, such as integer-autoregressive (INAR) and binomial-mixed Poisson INAR. We leave this as a task for our future study.

3.8 Appendix

The following Lemma is needed to derive the results of Theorems 3 and 4:

Lemma 1. Let $\varphi$ be a Borel measurable and real-valued integrable function and let $X \sim N(\theta, D)$, where $D$ is a positive semi-definite matrix with rank $l \leq n$. Also assume $V$ be symmetric and positive definite matrix such that $V^{1/2}DV^{1/2}$ is an idempotent matrix, and $VDV\theta = V\theta$. Then

1) $E(\varphi(Z^TVDVZ)Z) = \theta E(\varphi(\chi_{l+2}^2(\theta^T V \theta)))$

2) $E(\varphi(Z^TVDVZ)Z^T AZ) = E(\varphi(\chi_{l+2}^2(\theta^T V \theta))) \text{tr}(AD) + E(\varphi(\chi_{l+4}^2(\theta^T V \theta))) \theta^T A \theta,$

where $A$ is nonnegative definite matrix. The outline of the proof of this lemma is given in Nkurunziza (2016)[45].
Proof of Theorem 3

In this proof, we derive the bias expressions of the proposed estimators. It is obvious
that $\text{ADB}(\gamma) = 0$. The ADB of RMLE, SE, PSE, and IPT are as follows:

$$
\text{ADB}(\hat{\gamma}) = \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) \right) = -\Sigma R^T (R \Sigma R^T)^{-1} d = \vartheta
$$

$$
\text{ADB}(\hat{\gamma}_S) = \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma}_S - \gamma) \right) = - \lim_{n \to \infty} E \left( \sqrt{n} ((l_2 - 2) \hat{D}_n^{-1} (\hat{\gamma} - \gamma)) \right)
$$

$$
= -(l_2 - 2) \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) \hat{D}_n^{-1} \right) = (l_2 - 2) \vartheta E(Z_1^{-1})
$$

$$
\text{ADB}(\hat{\gamma}_{S+}) = \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma}_{S+} - \gamma) \right)
$$

$$
= \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma}_S - \gamma) - \sqrt{n} \left( 1 - (l_2 - 2) \hat{D}_n^{-1} \right) I \left( \hat{D}_n < (l_2 - 2) \right) (\hat{\gamma} - \gamma) \right)
$$

$$
= \text{ADB}(\hat{\gamma}_S) - \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) \left( 1 - (l_2 - 2) \hat{D}_n^{-1} \right) I \left( \hat{D}_n < (l_2 - 2) \right) \right)
$$

$$
= \text{ADB}(\hat{\gamma}_S) + \vartheta E \left( I \left( \hat{D}_n < (l_2 - 2) \right) \right) - \vartheta (l_2 - 2) E \left( \sqrt{n} (\hat{\gamma} - \gamma) \hat{D}_n^{-1} I \left( \hat{D}_n < (l_2 - 2) \right) \right)
$$

$$
= \text{ADB}(\hat{\gamma}_S) + \vartheta \Psi_{l_2+2}(l_2 - 2, \Delta) - (l_2 - 2) \vartheta E(Z_1^{-1} (Z_1 < (l_2 - 2)))
$$

$$
\text{ADB}(\hat{\gamma}_{\text{ipt}}) = \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma}_{\text{ipt}} - \gamma) \right)
$$

$$
= \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) \left( 1 - (l_2 - 2) \hat{D}_n^{-1} \right) I \left( \hat{D}_n > \chi_{l_2,\alpha}^2 \right) \right) + \text{ADB}(\hat{\gamma})
$$

$$
= \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) I \left( \hat{D}_n < \chi_{l_2,\alpha}^2 \right) \right)
$$

$$
- \lim_{n \to \infty} E \left( \sqrt{n} (\hat{\gamma} - \gamma) (l_2 - 2) \hat{D}_n^{-1} I \left( \hat{D}_n \leq \chi_{l_2,\alpha}^2 \right) \right) + \vartheta
$$

$$
= \lim_{n \to \infty} E \left( \sqrt{n} (l_2 - 2) \hat{D}_n^{-1} I \left( \hat{D}_n > \chi_{l_2,\alpha}^2 \right) \right) - \lim_{n \to \infty} E \left( \sqrt{n} (l_2 - 2) \hat{D}_n^{-1} (\hat{\gamma} - \gamma) \right)
$$

$$
+ \lim_{n \to \infty} E \left( \sqrt{n} (l_2 - 2) \hat{D}_n^{-1} I \left( \hat{D}_n \leq \chi_{l_2,\alpha}^2 \right) (\hat{\gamma} - \gamma) \right) + \vartheta
$$

$$
= \vartheta \left( 1 - \Psi_{l_2+2}(\chi_{l_2,\alpha}^2, \Delta) - (l_2 - 2) E \left( Z_1^{-1} (Z_1 \geq \chi_{l_2,\alpha}^2) \right) \right)$$
Proof of Theorem 4

Based on the definition of ADR function, it is necessary to derive the asymptotic covariance matrices for the four estimators. The covariance matrix of any estimator $\hat{\xi}^*$ is defined as:

$$\text{Cov}(\hat{\gamma}^*) = \lim_{n \to \infty} E\left(n(\hat{\gamma}^* - \gamma)(\hat{\gamma}^* - \gamma)^\top\right).$$

First, we will start deriving the covariance matrices of the UMPLE and RMPLE:

$$\text{Cov}(\hat{\gamma}) = \lim_{n \to \infty} E\left(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^\top\right) = G_{11}$$

$$\text{Cov}(\tilde{\gamma}) = \lim_{n \to \infty} E\left(\sqrt{n}(\tilde{\gamma} - \gamma)\sqrt{n}(\tilde{\gamma} - \gamma)^\top\right) = G_{12} + \vartheta\vartheta^\top.$$

Third, we derive the covariance matrices of the shrinkage and positive shrinkage estimators:

$$\text{Cov}(\hat{\gamma}_S) = \lim_{n \to \infty} E\left(\sqrt{n}(\hat{\gamma}_S - \gamma)\sqrt{n}(\hat{\gamma}_S - \gamma)^\top\right)$$

$$= \lim_{n \to \infty} E\left(\sqrt{n} \left(\hat{\gamma} - \gamma - (l_2 - 2)\hat{D}_n^{-1}(\hat{\gamma} - \tilde{\gamma})\right)\right) \sqrt{n} \left(\hat{\gamma} - \gamma - (l_2 - 2)\hat{D}_n^{-1}(\hat{\gamma} - \tilde{\gamma})\right)^\top$$

$$= E(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^\top) + (l_2 - 2)^2 \lim_{n \to \infty} E(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^\top \hat{D}_n^{-2})$$

$$- 2(l_2 - 2) \lim_{n \to \infty} E(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^\top \hat{D}_n^{-1})$$

$$= G_{11} + G_{13} E\left(Z_1^{-2}\right) + \vartheta\vartheta^\top E\left(Z_2^{-2}\right)$$

$$- 2(l_2 - 2) \lim_{n \to \infty} E(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^\top \hat{D}_n^{-1}).$$
Consider the last term:

\[
\lim_{n \to \infty} E(\sqrt{n}(\hat{\gamma} - \gamma) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})^T)
+ \hat{D}_n^{-1} = \lim_{n \to \infty} E\left( (\sqrt{n}(\hat{\gamma} - \gamma) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})^T \hat{D}_n^{-1} | \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \right)
\]

\[
= \lim_{n \to \infty} E\left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) (\sqrt{n}(\hat{\gamma} - \tilde{\gamma})^T - E(\sqrt{n}(\hat{\gamma} - \tilde{\gamma}))^T) \hat{D}_n^{-1} \right)
\]

\[
= \lim_{n \to \infty} E\left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})^T \hat{D}_n^{-1} \right) - \lim_{n \to \infty} E\left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \hat{D}_n^{-1} \right) E\left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \right)^T
\]

\[
= \mathcal{G}_{13} E\left( \chi_{l_2+2}^{-2}(\Delta) \right) + \vartheta \vartheta^T E\left( \chi_{l_2+4}^{-2}(\Delta) \right) - \vartheta \vartheta^T E\left( \chi_{l_2+2}^{-2}(\Delta) \right)
\]

\[
= \mathcal{G}_{13} E\left( Z_1^{-1} \right) + \vartheta \vartheta^T E\left( Z_2^{-1} \right) - \vartheta \vartheta^T E\left( Z_1^{-1} \right)
\]

Hence

\[
\text{Cov}(\hat{\gamma}_S) = \mathcal{G}_{11} + (l_2 - 2)^2 \left( \mathcal{G}_{13} E\left( Z_1^{-2} \right) + \vartheta \vartheta^T E\left( Z_2^{-2} \right) \right)
- 2(l_2 - 2) \left( \mathcal{G}_{13} E\left( Z_1^{-1} \right) + \vartheta \vartheta^T E\left( Z_2^{-1} \right) - \vartheta \vartheta^T E\left( Z_1^{-1} \right) \right)
= \mathcal{G}_{11} + ((l_2 - 2)^2 E\left( Z_1^{-2} \right) - 2(l_2 - 2) E\left( Z_1^{-1} \right)) \mathcal{G}_{13}
+ \left( (l_2 - 2)^2 E\left( Z_2^{-2} \right) + 2(l_2 - 2) E\left( Z_1^{-1} \right) - 2(l_2 - 2) E\left( Z_2^{-1} \right) \right) \vartheta \vartheta^T
\]
Let $F_m(\Delta) = \left(1 - (l_2 - 2)\hat{D}_n^{-1}\right)^m I \left(\hat{D}_n < (l_2 - 2)\right)$, where $m = 1, 2$

\[
\text{Cov}(\hat{\gamma}_S+) = \lim_{n \to \infty} E \left(\sqrt{n}(\hat{\gamma}_S+ - \gamma)\sqrt{n}(\hat{\gamma}_S+ - \gamma)^\top\right),
\]

\[
= \lim_{n \to \infty} E \left(\sqrt{n}(\hat{\gamma}_S - \gamma)\sqrt{n}(\hat{\gamma}_S - \gamma)^\top\right)
+ \lim_{n \to \infty} E \left(F_2(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma} - \bar{\gamma})^\top\right)
- 2 \lim_{n \to \infty} E \left(F_1(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma}_S - \gamma)^\top\right)
= \text{Cov}(\hat{\gamma}_S) + \lim_{n \to \infty} E \left(F_2(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma} - \bar{\gamma})^\top\right)
- 2 \lim_{n \to \infty} E \left(F_1(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\left(\sqrt{n}(\hat{\gamma} - \gamma)^\top + \left(1 - (l_2 - 2)\hat{D}_n^{-1}\right)\sqrt{n}(\hat{\gamma} - \bar{\gamma})^\top\right)\right)
= \text{Cov}(\hat{\gamma}_S) - \lim_{n \to \infty} E \left(F_2(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma} - \bar{\gamma})^\top\right)
- 2 \lim_{n \to \infty} E \left(F_1(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma} - \gamma)^\top\right).
\]

Consider the second term:

\[- \lim_{n \to \infty} E \left(F_2(\Delta)\sqrt{n}(\hat{\gamma} - \bar{\gamma})\sqrt{n}(\hat{\gamma} - \bar{\gamma})^\top\right)
= - \lim_{n \to \infty} E \left(\left(1 - (l_2 - 2)\hat{D}_n^{-1}\right)^2 I \left(\hat{D}_n < (l_2 - 2)\right)\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\bar{\gamma} - \bar{\gamma})^\top\right)
= - \mathcal{G}_{13} E \left(I (Z_1 < (l_2 - 2)) (1 - (l_2 - 2)Z_1^{-1})^2\right)
- \vartheta^\top E \left(I (Z_2 < (l_2 - 2)) (1 - (l_2 - 2)Z_2^{-1})^2\right).\]
Consider the third term:

\[
-2 \lim_{n \to \infty} E \left( F_1(\Delta) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})(\sqrt{n}(\hat{\gamma} - \gamma)^T) \right) \\
= -2 \lim_{n \to \infty} E \left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma})E \left( F_1(\Delta) \sqrt{n}(\hat{\gamma} - \gamma)^T | \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \right) \right) \\
= -2 \lim_{n \to \infty} E \left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma})E \left( F_1(\Delta) \sqrt{n}(\hat{\gamma} - \gamma)^T \right) F_1(\Delta) + 0 \right) \\
= -2 \lim_{n \to \infty} E(\sqrt{n}(\hat{\gamma} - \tilde{\gamma}))I(\hat{D}_n < (l_2 - 2)) - (l_2 - 2)\hat{D}_n^{-1}\sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \\
- (\hat{\gamma})I(\hat{D}_n < (l_2 - 2))E(\sqrt{n}(\hat{\gamma} - \gamma)^T) \\
= 2\Psi_{l_2+2}(l_2 - 2, \Delta)\vartheta\vartheta^T - 2(l_2 - 2)E \left( Z_1^{-1}I \left( Z_1 < (l_2 - 2) \right) \right) \vartheta\vartheta^T \\
= (2\Psi_{l_2+2}(l_2 - 2, \Delta) - 2(l_2 - 2)E \left( Z_1^{-1}I \left( Z_1 < (l_2 - 2) \right) \right)) \vartheta\vartheta^T.
\]

Finally,

\[
\text{Cov} (\hat{\gamma}_{S+}) = \text{Cov} (\hat{\gamma}_S) - E \left( (1 - (l_2 - 2)Z_1^{-1})^2 I \left( Z_1 < (l_2 - 2) \right) \right) \mathcal{G}_{13} \\
+ (2\Psi_{l_2+2}(l_2 - 2, \Delta) - 2(l_2 - 2)E \left( Z_1^{-1}I \left( Z_1 < (l_2 - 2) \right) \right)) \vartheta\vartheta^T \\
- E \left( (1 - (l_2 - 2)Z_2^{-1})^2 I \left( Z_2 < (l_2 - 2) \right) \right) \vartheta\vartheta^T.
\]
Four, we derive the covariance matrices of the improved pretest estimator:

\[
\text{Cov}(\hat{\gamma}_{ipt}) = \lim_{n \to \infty} \mathbb{E} \left( \sqrt{n} (\hat{\gamma}_{ipt} - \gamma) \sqrt{n} (\hat{\gamma}_{ipt} - \gamma)^\top \right) \\
= \lim_{n \to \infty} \mathbb{E} \left( \sqrt{n} (\hat{\gamma}_S - \gamma - (\hat{\gamma} - \tilde{\gamma})(1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) ) \right) \\
\times \lim_{n \to \infty} \mathbb{E} \left( \sqrt{n} (\hat{\gamma}_S - \gamma - (\hat{\gamma} - \tilde{\gamma})(1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) ) \right)^\top \\
= \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma}_S - \gamma)(\hat{\gamma}_S - \gamma)^\top \right) \\
+ \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})(\hat{\gamma} - \tilde{\gamma})^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right) \\
- 2 \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})(\gamma - \gamma)^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right) \\
= \text{ADR}(\hat{\gamma}_S) + \mathcal{G}_{13} \mathbb{E} \left( (1 - (l_2 - 2)Z_1^{-1})^2 I(Z_1 \leq \chi_{l_2, \alpha}^2) \right) \\
+ \theta\theta^\top \mathbb{E} \left( (1 - (l_2 - 2)Z_2^{-1})^2 I(Z_2 \leq \chi_{l_2, \alpha}^2) \right) \\
- 2 \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})(\hat{\gamma}_S - \gamma)^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right).
\]

Consider the fourth term:

\[
= -2 \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})(\hat{\gamma} - \gamma)^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right) \\
- 2 \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})(\hat{\gamma} - \tilde{\gamma})^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right) \\
= -2 \lim_{n \to \infty} \mathbb{E} \left( (\hat{\gamma} - \tilde{\gamma})E(\hat{\gamma} - \gamma)^\top (1 - (l_2 - 1)\hat{D}_n^{-1})I(\hat{D}_n \leq \chi_{l_2, \alpha}^2) \right) \\
- 2\mathcal{G}_{13} \mathbb{E} \left( (1 - (l_2 - 2)Z_1^{-1})I(Z_1 \leq \chi_{l_2, \alpha}^2) \right) - 2\theta\theta^\top \mathbb{E} \left( (1 - (l_2 - 2)Z_2^{-1})^2 I(Z_2 \leq \chi_{l_2, \alpha}^2) \right) \\
= -2\theta\theta^\top \mathbb{E} \left( Z_1^{-1}I(Z_1 \leq \chi_{l_2, \alpha}^2) \right) - 2\mathcal{G}_{13} \mathbb{E} \left( (1 - (l_2 - 2)Z_1^{-1})I(Z_1 \leq \chi_{l_2, \alpha}^2) \right) \\
- 2\theta\theta^\top \mathbb{E} \left( (1 - (l_2 - 2)Z_2^{-1})^2 I(Z_2 \leq \chi_{l_2, \alpha}^2) \right).
\]

The ADR expressions in Theorem 4 now follow from equation(16) which completes
the proof.
Chapter 4

Efficient estimation method for generalized ARFIMA models

Abstract

This paper focuses on pretest and shrinkage estimation strategies for generalized autoregressive fractionally integrated moving average (GARFIMA) models when some of the regression parameters are possible to restrict to a subspace. These estimation strategies are constructed on the assumption that some covariates are not statistically significant for the response. To define the pretest and shrinkage estimators, we fit two models: one includes all the covariates and the others are subject to linear constraint based on the auxiliary information of the insignificant covariates. The unrestricted and restricted estimators are then combined optimally to get the pretest and shrinkage estimators. We enlighten the statistical properties of the shrinkage and pretest estimators in terms of asymptotic bias and risk. We
examine the comparative performance of pretest and shrinkage estimators with respect to
unrestricted maximum partial likelihood estimator (UMPLE). We show that the shrinkage
estimators have a lower relative mean squared error as compared to the UMLE, when the
number of significant covariates exceeds two. Monte Carlo simulations are conducted to
examine the relative performance of the proposed estimators to the UMLE. An empirical
application is used for the usefulness of our proposed estimation strategies.

\textit{Keywords:} ARFIMA; asymptotic distributional bias and risk; Monte Carlo simulation;
partial likelihood; shrinkage and pretest estimators

4.1 Introduction

Time series analysis and modelling are scientific ways to predict the future. Time
series technique is extensively applied in a financial and business forecast based on
the historical pattern of data collected over time and comparing it with the current
trend of data. For example, the change in stock price depends on many economic
factors and it can be forecasted through time series analysis. There are many types
of time series models that bring out the desired results. Among the existing models,
a fundamental one is the autoregressive moving average (ARMA) model, originated
from the autoregressive model (AR) and the moving average model (MA). These
models along with the autoregressive integrated moving average model (ARIMA),
are useful to capture short-range dependence, as they relate to the traditional non-fractional order models Sheng (2011)[61]. A drawback of these models may occur for the case of non-stationarity in long memory high frequency time series data. The autoregressive fractionally integrated moving average model, introduced by Granger (1980)[26] and Honsking (1981)[29], may play an important role for modeling long memory high-frequency time series data in the field of economics, medicine, biology, hydrology, astronomy, and many other areas of research Contreras (2013)[12].

In the past, many researchers suggested various approaches for fitting ARMA, βARMA, ARFIMA, and βARFIMA processes. Pumi (2018)[54] introduced a class of beta ARFIMA models for continuous random variables taking values in the continuous unit interval (0, 1). They accommodated a set of regressors and a long-range dependent time series structure in the model. Rocha (2009)[57] proposed βARMA model that includes both autoregressive and moving average dynamics, and also includes a set of regressors. Tsai (2009)[66] compared the continuous-time ARFIMA model with discrete-time ARFIMA for examining time-series data of short and long memory dependence. The remarkable example of ARFIMA, in the R-package mentioned by Palma (2007)[49], implemented statistical tools for analyzing ARFIMA in terms of predictive capability, exact autocovariance, and parameter estimation. Fokianos (2004)[21] showed that models for time series of counts fall within the broad
class of time series following generalized linear models, and their analysis is based on partial likelihood inference. Doornik (2003)[15] implemented the improvement of computational properties of ARFIMA models by computing the quasi-likelihood exactly, using the autocovariance functions up to order $n$. Reisen(2001)[56] carried out an extensive simulation study comparing both the semiparametric and parametric approaches in ARFIMA processes. Li (1994)[38] proposed the problem of extending the classical moving average models to time series with conditional distributions given by generalized linear models.

This paper considers the problem of extending ARFIMA time series models to a non-Gaussian framework, called GARPIMA and then we apply the James-Stein shrinkage and pretest estimation procedure for GARPIMA model when there are many covariates subject of the investigation. Optimal selection of active covariates is a challenging task for researchers in the model due to over-modeling. They try to reduce the number of passive variables in the model. The main aim is to optimize the predictive capability while reducing the number of covariates in the model. The use of James-Stein shrinkage and pretest estimation enables us to carry out the above goal by using the auxiliary information based on the inactive covariates. Notably, the regression parameter vector $\gamma$ of dimension $l \times 1$ divided into dual sub-vectors $\gamma = (\gamma_1^T, \gamma_2^T)^T$ with size $l_1 \times 1$ of $\gamma_1$ and $l_2 \times 1$ of $\gamma_2$ respectively, such that $l = l_1 + l_2$. Here $\gamma_1$ and $\gamma_2$ represented main and nuisance effects, respectively. Our aim in this
paper to estimate $\gamma_1$ using the information from nuisance effects vector $\gamma_2$ when values of $\gamma_1$ are close to some prescribed values along with the hypothesis $H_0 : \gamma_2 = 0$. This hypothesis can be written in general form $H_0 : R\gamma = 0$, where $R = [O, I]$. Here $O$ and $I$ are zeros and identity matrices of orders $l_2 \times l_1$ and $l_2 \times l_2$ respectively. Hossain (2016)[32] used this parameter partition for time series following GLM and applied the pretest and shrinkage techniques to estimate the mean response parameters. In the proposed work, we extend these approaches for GARFIMA model.

There has recently been considerable attention on applying James-Stein shrinkage ideas to parameter estimation in linear models with AR, ARMA, and GRACH errors, and time series following generalized linear models. The application of shrinkage estimator for linear models with ARMA and GARCH errors can be found in Paolella (2019)[52]. Yoon (2017)[74] focused on adaptive LASSO method for the linear regression models with the ARMA errors. Hossain (2016)[32] considered the James–Stein shrinkage and pretest estimation methods for time series following generalized linear models when some of the regression parameters may be restricted to a subspace. Hossain (2014)[30] proposed the non-penalty shrinkage estimation for linear regression model with autoregressive errors. Fallahpour(2014)[19] proposed a shrinkage estimation strategy and implement variable selection methods, such as lasso and adaptive lasso strategies for multiple regression models with first-order random
The paper is structured as follows. In Section 4.2 we present the model specification of GARFIMA. In Section 4.3 we set up a parameter estimation, the partial score vector and conditional information matrix. Section 4.4 examines the asymptotic distributional bias and risk of the proposed estimators. In Section 4.5, we introduce a Monte Carlo simulation study to evaluate the numerical performance of the maximum partial likelihood estimator. An application of GARFIMA models to financial and business data is represented in Section 4.6 Section 4.7 closes the article with concluding remarks.

4.2 Model specification of GARFIMA

Let \( \{y_t\}_{t=1}^{\infty} \) be a process of interest and focusing towards prediction and the use of maximum partial likelihood inference. Let \( \mathbf{x}_{t-1}^\top \) denote the \( l \)-dimensional vector of covariates at time \( t - 1 \) and these covariates up to time \( t \) are considered to be included in the model. Let \( \mathcal{F}_{t-1} \) denote the \( \sigma \)-field generated by the past and present covariates and possibly past values of the response variable. In the generalized ARMA (GARMA) model, the conditional distribution of each observation \( y_t \), for \( t = 1, 2, \ldots, n \) given the past information \( \mathcal{F}_{t-1} = \{y_{t-1}, \ldots, y_1, \mathbf{x}_t, \cdots \mathbf{x}_1\} \) is assumed
to belong to the exponential family. The density of this family is:

\[
f (y_t; \theta_t, \psi | \mathcal{F}_{t-1}) = \exp \left( \frac{y_t \theta_t - b(\theta_t)}{a_t(\psi)} + c(y_t; \psi) \right), \quad t = 1, 2, \cdots, n. \tag{4.1}
\]

where \(a_t(\cdot), b(\cdot),\) and \(c(\cdot)\) are specific functions that define the particular exponential family, \(\psi\) is a dispersion parameter, and \(\theta_t\) is the natural parameter. Following McCullagh (1989)[40], we get \(\mu_t = \mathbb{E}(y_t|\mathcal{F}_{t-1}) = b'(\theta_t)\) and \(\sigma^2_t = \text{Var}(y_t|\mathcal{F}_{t-1}) = a_t(\psi) b''(\theta_t)\) along the usual notation of derivatives.

Unlike the classical GLM, the linear predictor is \(\eta_t = x_{t-1}^\top \gamma + \tau_t\), where \(\gamma = (\gamma_1, \cdots, \gamma_l)^\top\). Here, the additional component, \(\tau_t\), allows ARMA model terms to be included additively in the predictor. The past values of time-dependent covariates, \(x_t\), are also included in the predictor. Thus the general model for \(\mu_t\) is given by \(g(\mu_t) = \eta_t = x_{t-1}^\top \gamma + \tau_t\), with:

\[
\tau_t = \sum_{j=1}^{p} \phi_j (g(y_{t-j}) - x_{t-j-1}^\top \gamma) + \rho_t \sum_{j=1}^{q} \delta_j \rho_{t-j}, \tag{4.2}
\]

where \(\rho_{t-j} = g(y_{t-j}) - \eta_{t-j}\) and \(\phi = (\phi_1, \phi_2, \cdots, \phi_p)^\top\), and \(\delta = (\delta_1, \delta_2, \cdots, \delta_q)^\top\) are the autoregressive and moving average parameters, respectively. The function \(g(\cdot)\) is a twice differentiable monotonic one-to-one link function, while \(\eta_t\) is known as linear predictor of the model. Model (4.2) has been termed as generalized linear autoregressive moving average Fokianos (2004)[21].
The statistical literature has been concerned with the study of long-range dependence models that go beyond the presence of random walks and unit roots in the univariate time series processes. The ARFIMA process is a class of long-range models Granger (1980)[26], Honsking(1981) [29]. The main objective of this model is to explicitly account for persistence to incorporate the long-term correlations in the data. We consider a process \( \{x_t\} \) that follows ARFIMA\((p,d,q)\) when \( \{x_t\} \) is a stationary solution of:

\[
\phi(L)(1 - L)^d x_t = \delta(L)\varepsilon_t,
\]

(4.3)

where \( L \) denotes the backward shift operator \( L^k(x_t) = x_{t-k} \) for \( k = 1, 2, 3, \ldots \) and \( \varepsilon_t \) represents the white noise. However, \( \phi(x) = -\sum_{i=0}^{p} \phi_i x^i \) and \( \delta(x) = \sum_{j=0}^{q} \delta_j x^j \) are the AR\((p)\) and MA\((q)\) processes, respectively, and \( \phi(x) \) and \( \delta(x) \) have no common roots. The fractional part of ARFIMA\((p,d,q)\) is represented by \((1 - L)^{-d} = \sum_{k=0}^{\infty} \pi_k L^k \) where \( \pi_k = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)} \), \( k \geq 1 \) along with \( \pi_0 = 1 \) and \( d \in (-0.5, 0.5) \). The fractional integration parameter \( d \) is used to capture the long-run effects, that is, an indication that this process might be useful is a slowly tapering sample autocorrelation without particularly high autocorrelations. More detailed of ARFIMA\((p,d,q)\) model can be found in Palma(2007)[49]. Therefore, the generalization of GARMA\((p,q)\) in (4.2) can be expressed as:

\[
\eta_t = \eta_{t-1} + \gamma + \sum_{j=1}^{p} \phi_j (y_{t-j} - \eta_{t-1-\gamma} + \sum_{k=0}^{\infty} c_k \rho_{t-k},
\]

(4.4)

where \( \delta_k = 0 \), for \( k > q \), \( c_0 = 1 \), \( c_k = \sum_{i=0}^{\min(k,q)} \delta_i \pi_{k-i}, k > 0 \). More details of the
model can be found in Pumi(2018) [54]. The error term is defined in a recursive fashion in the prediction scale, that is, we consider \( \rho_t = g(y_t) - \eta_t \). Models (4.1) along with (4.4) can be considered as GARFIMA models.

4.3 Estimation of GARFIMA models

The estimation of GARFIMA model, described herein performs conditional partial maximum likelihood method. Let \( \{(y_t, x_t^T)\} \) be a sample from proposed model for \( t = 1, 2, 3, \ldots, n \) and let \( \beta = (\gamma^T, \xi^T)^T \) with \( \xi^T = (\phi^T, \delta^T, d)^T \) be \( (l + p + q + 1) \) dimensional parameters vector to be estimated. These parameters are estimated by maximum likelihood method. Since the covariates, \( x_t \), are stochastic, the estimation of \( \beta \) is done by maximizing the partial likelihood Fahrmeir (2001)[18, pg. 194]. The partial likelihood (PL) function is defined by:

\[
\text{PL}(\beta) = \prod_{t=1}^{n} f(y_t; \theta_t, \varphi|F_{t-1}).
\]

The partial log-likelihood function \( \ell(\beta) = \log \text{PL}(\beta) \) can be written as:

\[
\ell(\beta) = \sum_{t=1}^{n} f(y_t; \theta_t, \varphi|F_{t-1}) = \sum_{t=1}^{n} \left( \frac{y_t u(\eta_t) - b(u(\eta_t))}{a_t(\varphi)} + c(y_t; \varphi) \right) = \sum_{t=1}^{n} l_t,
\]

where \( l_t = (y_t u(\eta_t) - b(u(\eta_t)))/a_t(\varphi) + c(y_t; \varphi) \), \( \eta_t = x_{t-1}^T \gamma \). In order to emphasize the dependence of the partial likelihood on \( \gamma \), we have used the notation \( u(\cdot) \equiv (g \circ \mu(\cdot))^{-1} = \mu^{-1}(g^{-1}(\cdot)) \), where the small circle \( \circ \) represents composition of functions.
and $\mu$ so that $\theta_t = u(x_{t-1}^\top \gamma)$. The partial log-likelihood function is $\ell(\beta) = \sum_{t=1}^{n} l_t$ and henceforth, the partial maximum likelihood estimator of $\beta$ is defined as:

$$\hat{\beta} = \arg\max_{\beta \in \Omega} \ell(\beta),$$

where $\Omega \subseteq \mathbb{R}^{l+p+q} \times (-0.5, 0.5)$ be the parameter space and we call $\hat{\beta}$ the unrestricted maximum partial likelihood estimator (UMPLE). In the next subsections we derive the score vector related to the maximization. Initially, we were interested in estimating all the parameters but our primary interest is on the regression parameter vector $\gamma$, while the other ARFIMA model parameter vector $\xi^\top = (\phi^\top, \delta^\top, d)^\top$ serves as a nuisance.

### 4.3.1 Partial Score Vector

To develop the conditional score vector, we obtain the derivative of $\ell(\beta)$ given in (4.5) with respect to $\beta = (\gamma^\top, \phi^\top, \delta^\top, d)^\top$. More information about the conditional score vector can be found in Pumi (2018)[54] and Kedem(2002)[36, chap. 2]. Now the derivative of the log-likelihood is:

$$\frac{\partial \ell(\beta)}{\partial \beta_j} = \sum_{t=1}^{n} \frac{\partial l_t}{\partial \theta_t} \cdot \frac{\partial \theta_t}{\partial \mu_t} \cdot \frac{\partial \mu_t}{\partial \eta_t} \cdot \frac{\partial \eta_t}{\partial \beta_j} = \sum_{t=1}^{n} \frac{y_t - \mu_t(\beta)}{\sigma_t^2(\beta)} \cdot \frac{\partial \mu_t}{\partial \eta_t} \cdot \frac{\partial \eta_t}{\partial \beta_j}, \quad (4.6)$$

The derivatives of $\ell(\beta)$ can be simplified to determining the derivatives $\frac{\partial \mu_t}{\partial \beta_j}$ for each component of the parameter vector $\beta$. The similar expressions of these derivatives can be found in Pumi (2018)[54]. Let $A$, $B$ and $C$ be the matrices of order $n \times l$,
\( n \times p, n \times q \), respectively whose \((s, t)th\) elements are given by 
\[
A_{t,s} = \frac{\partial \eta}{\partial \gamma_s}, \quad B_{t,s} = \frac{\partial \eta}{\partial \phi_s}, \quad C_{t,s} = \frac{\partial \eta}{\partial \delta_s},
\]
Let 
\[
y = (y_1 - \mu_1, y_2 - \mu_2, \cdots, y_n - \mu_n)^\top, \quad \zeta = \left( \frac{\partial \eta_1}{\partial d}, \frac{\partial \eta_2}{\partial d}, \cdots, \frac{\partial \eta_n}{\partial d} \right)^\top \text{ and } T = \text{diag}(1/\sigma_1^2(\beta), 1/\sigma_2^2(\beta), \cdots, 1/\sigma_n^2(\beta)).
\]
Hence, the matrix form of score vector can be mentioned as 
\[
W(\beta) = (W_\gamma(\beta)^\top, W_\phi(\beta)^\top, W_\delta(\beta)^\top, W_d(\beta))^\top \in \mathbb{R}^{l+p+q+1}.
\]
Here, 
\[
W_d(\beta) = \zeta^\top Ty, \quad W_\gamma(\beta) = A^\top Ty, \quad W_\phi(\beta) = B^\top Ty, \quad W_\delta(\beta) = C^\top Ty.
\]
We obtained the conditional maximum likelihood estimates by solving the non-linear system 
\[
W(\beta) = 0,
\]
where 0 shows the null vector in \(\mathbb{R}^{l+p+q+1}\). The solution of this non-linear system is obtained iteratively by the method of Fisher scoring, a modification of the well known Newton-Raphson method.

### 4.3.2 Conditional Information Matrix

The Fisher conditional information matrix plays an important role in driving the asymptotic properties of the partial maximum likelihood estimator. More information and elements of Fisher conditional information matrix can be found in [54] and Kedem (2002)[36, chap. 2].

\[
E \left( \frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^\top} \middle| \mathcal{F}_{t-1} \right) = - \sum_{t=1}^{n} w_t \cdot \frac{\partial \eta_t}{\partial \beta_t} \frac{\partial \eta_t}{\partial \beta_t^\top},
\]
where 
\[
w_t = \left( \frac{\partial \eta_t}{\partial \eta_t} \right)^2 (\sigma_t^2(\beta))^{-1}
\]
In general, for each co-ordinate \(\beta_j\) of \(\beta = (\gamma^\top, \phi^\top, \delta^\top, d)^\top\) for all \(i, j \in (1, \cdots, l + p + q + 1)\) such that 
\[
G_{\beta_i, \beta_j}(\beta) = - E \left( \frac{\partial^2 \ell(\beta)}{\partial \beta_i \partial \beta_j} \middle| \mathcal{F}_{t-1} \right),
\]
for \(i \leq j\) and \(G_{\beta_i, \beta_j}(\beta) = G_{\beta_j, \beta_i}(\beta)^\top\), for \(i > j\). Therefore, the information matrix is the matrix \(G_n(\beta)\) whose \((i,j)\) element is \(G_{\beta_i, \beta_j}(\beta)\). Next, we state the regularity conditions for the asymptotic distribution of UMPLE, \(\hat{\beta}\) in Theorem 1.

**Regularity Conditions Kedem(2002)[see 36, chap. 1]:**

1. The true parameter \(\beta\) belongs to an open set \(\Omega \subseteq \mathbb{R}^{l+p+q+1}\).
2. The covariate vector \(x_{t-1}\) almost surely lies in compact set \(\Gamma \subseteq \Omega\), such that \(Pr(\sum_{t=1}^{n} x_{t-1} x_{t-1}^\top > 0) = 1\). Also, \(x_{t-1} \gamma\) lies almost surely in the domain of the inverse link function \(h = g^{-1}\) for all \(x_{t-1} \in \Gamma\) and \(\beta \in \Omega\).
3. The inverse link function \(h\) defined in (2) is twice continuously differentiable with respect to \(\beta\).
4. There exists a probability measure \(\varrho\) in \(\Omega\) such that \(\int_{\Omega} x_{t-1} x_{t-1}^\top \varrho(dx_{t-1})\) is positive definite and such that under (1) and (2) for Borel sets \(A \subset \Omega\), \(1/n \sum_{t=1}^{n} I_{[x_{t-1} \in A]} \to \varrho(A)\) in probability as \(n \to \infty\), at the true value of \(\beta\).

Conditions (1) and (3) guarantee that the second derivative of the log-partial likelihood is continuous with respect to \(\beta\). In addition, the condition \(\partial h(x)/\partial x \neq 0\), \(\forall x \in \mathbb{R}\) together with (2), assuming \(n\) is large, implies that the conditional information matrix is positive definite with probability 1.

**Theorem 1:** Under some regularity conditions in above, \(\hat{\beta}\) is almost surely unique
for sufficiently large $n$,

i. \( \hat{\beta} \to \beta \) in probability as \( n \to \infty \),

ii. \( \sqrt{n}(\hat{\beta} - \beta) \overset{\mathcal{L}}{\to} N(0, G^{-1}(\beta)) \) as \( n \to \infty \),

iii. \( \sqrt{n}(\hat{\beta} - \beta) - \frac{1}{\sqrt{n}} G^{-1}(\beta) \frac{\partial \ell(\beta)}{\partial \beta} \to 0 \) as \( n \to \infty \), where \( G(\beta) = \lim_{n \to \infty} \frac{G_n(\beta)}{n} \).

**Proof**: See Appendix.

Assume \( \beta = (\gamma^T, \xi^T)^T \), we can partition the information matrix as:

\[
G_n(\beta) = \begin{bmatrix}
    V_{\gamma \gamma} & V_{\gamma \xi} \\
    V_{\xi \gamma} & V_{\xi \xi}
\end{bmatrix},
\]

(4.7)

where \( V_{\gamma \gamma} \) and \( V_{\xi \xi} \) are positive-definite matrices of size \( l \times l \) and \( (p+q+1) \times (p+q+1) \), respectively.

### 4.3.3 The restricted maximum partial likelihood estimator

In many practical settings of time series data modelling, the regression parameters \( \gamma \) in (4.4), may be subject to restrictions \( R\gamma = 0 \), where \( R \) is a \( l_2 \times l \) matrix of rank \( l_2 \leq l \), and the other parameter vector \( \xi \) is treated as nuisance. These restrictions typically reflect auxiliary information about the value of the regression parameters. We use these restrictions to construct pretest and shrinkage estimators which will be defined in the next section and thereby expect estimation efficiency. Finding
the restricted estimator using the restrictions is more complicated compared to the unrestricted estimator.

We incorporate the restrictions by assuming the hypothesis \( H_0 : R\gamma = 0 \). Now, maximizing \( \ell(\beta) = \sum_{t=1}^{n} f(y_t; \theta_t, \varphi|F_{t-1}) \) subject to \( R\gamma = 0 \) is equivalent to maximizing

\[
\begin{align*}
\tilde{\beta} &= (\tilde{\gamma}^\top, \tilde{\xi}^\top)^\top = \arg\max_{\gamma, \xi} \left\{ \sum_{t=1}^{n} f(y_t; \theta_t, \varphi|F_{t-1}) : R\gamma = 0 \right\},
\end{align*}
\]

(4.8)

where \( \tilde{\beta} \) is the restricted maximum partial likelihood estimator (RMPLE). The solution above is again obtained iteratively by the method of Fisher scoring. In the simulation study, we will show that if the restriction is true, the RMPLE is expected to perform better than the UMPLE in terms of lower mean squared error sense, otherwise the RMPLE tends to be biased, inefficient, and even inconsistent.

### 4.3.4 The pretest and shrinkage estimators

The pretest and shrinkage estimators are based on the log-partial likelihood ratio test statistic \( D_n \). If \( \hat{\gamma} \) and \( \tilde{\gamma} \) are the UMPLE and RMPLE of \( \gamma \), then the test statistic for testing \( H_0 : R\gamma = 0 \) vs. \( H_0 : R\gamma \neq 0 \) is:

\[
D_n = 2 \left[ \ell(\hat{\gamma}; y_1, y_2, \cdots, y_n) - \ell(\tilde{\gamma}; y_1, y_2, \cdots, y_n) \right]
\]

Under \( H_0 \), the test statistic, \( D_n \), asymptotically follows a \( \chi^2 \)-distribution with \( l_2 \) degrees of freedom.
The pretest estimator (PT) of $\gamma$ based on $\hat{\gamma}$ and $\tilde{\gamma}$, will be defined as:

$$\hat{\gamma}_P = \hat{\gamma} - I(D_n \leq \chi^2_{l_2,\alpha})(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 1,$$

where $I(F)$ is an indicator function of a set $F$, and $\chi^2_{l_2,\alpha}$ is the $\alpha$-level critical value of the distribution with $l_2$ degrees of freedom. Based on the indicator function, $\hat{\gamma}_P$ chooses $\hat{\gamma}$ or $\tilde{\gamma}$ depending on whether $H_0$ is accepted or rejected, respectively. It is important to note that $\hat{\gamma}_P$ is bounded and performs better than $\tilde{\gamma}$ in some part of the parameter space. For details, see Judge (1978)[35], Nkurunziza (2013)[43] and Ahmed(2006)[2].

A limitation of the PT estimator is that it is a discontinuous function of $\hat{\gamma}$ and $\tilde{\gamma}$, and is dependent on the choice of $\alpha$ through the critical value $\chi^2_{l_2,\alpha}$. We address this limitation by defining the shrinkage estimator (SE) of $\gamma$ as:

$$\tilde{\gamma}_S = \hat{\gamma} + (1 - (l_2 - 2)D_n^{-1})(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 3.$$

The SE is the convex linear combination $\lambda\hat{\gamma} + (1 - \lambda)\tilde{\gamma}$, where $0 \leq \lambda \leq 1$ is the parameter that determines the extent to which these estimates are pooled together and sum of the coefficients is 1. If $\lambda = 0$, then the RMPLE dominates completely, whereas for $\lambda = 1$, then no shrinkage occurs. A setback with the SE is that it is not a convex combination of the UMPLE and RMPLE. This has the tendency to cause over-shrinking which gives the estimator the opposite sign of the UMPLE. To
overcome this issue, we use the positive-part shrinkage estimator (PSE)

\[
\hat{\gamma}_{s+} = \tilde{\gamma} + (1 - (l_2 - 2)D_n^{-1})^+(\hat{\gamma} - \tilde{\gamma}), \quad l_2 \geq 3,
\]

and where \( z^+ = \max(0, z) \).

**Remark.** Instead of using the pretest and shrinkage estimators, one can use the penalized methods for GARFIMA model. The commonality between the penalized methods Tibshirani(1996)[65], Efron(2004)[17], Friedman(2010)[22] and the pretest and shrinkage methods is that they both set some \( \gamma \) coefficients to zero for some data sets. These both methods also shrink maximum likelihood estimates. The discrepancy between two methods is that pretest and shrinkage methods select a linear restriction based on the prior research or the auxiliary information of the covariates but the penalized method treats all the covariates coefficients equally and it does not single out a linear restriction. We can use the pretest and shrinkage estimators over the penalized estimator, as they are free from tuning parameters and are easy to compute. One of the benefits of the shrinkage estimators is that it reduces the mean squared error by trading off some amount of bias for a reduction in variance.

In Section 4.4, we will investigate the properties of these methods when the restriction is correctly specified, the pretest and shrinkage methods would be expected to do well, and when the restriction is incorrectly specified and the penalty methods would be expected to do better.
4.4 Asymptotic results

In this section, we derive the asymptotic distributional bias (ADB) and asymptotic distributional risk (ADR) of the proposed estimators using the local asymptotic normality approach of Van (1998)[67]. The test statistic, $D_n$, converges to $\infty$ as $n \to \infty$ for the fixed alternative $H_a : R\gamma \neq 0 + c$, where $c = (c_1, c_2, \cdots, c_{l_2}) \in \mathbb{R}^{l_2}$ is a real fixed vector. Let $\hat{\gamma}^*$ be a generic notation for any of $\hat{\gamma}, \tilde{\gamma}, \hat{\gamma}_P, \hat{\gamma}_S$, or $\hat{\gamma}_{S^+}$. Under the fixed alternative hypothesis, it can be seen that $\sqrt{n}(\gamma^* - \gamma) = \sqrt{n}(\hat{\gamma} - \gamma) + o_p(1)$. That is, all of our estimators defined above are asymptotically equivalent to $\hat{\gamma}$. This implies that the asymptotic risk functions are all the same if $H_0$ is true. Thus, we cannot infer any difference among these estimators. To obtain the meaningful comparisons among the estimators, we consider a sequence of local alternatives

$$K_n : R\gamma = \frac{c}{\sqrt{n}}. \quad (4.9)$$

Under local $K_n$, the following theorem facilitates the derivation and numerical computation of the ADB and the ADR of the estimators outlined below:

**Theorem 2:** Under the local alternatives (4.9) and the regularity conditions in Subsection 3.2, the joint distribution is:

$$
\begin{bmatrix}
\sqrt{n}(\hat{\gamma} - \gamma) \\
\sqrt{n}(\tilde{\gamma} - \gamma) \\
\sqrt{n}(\hat{\gamma} - \tilde{\gamma})
\end{bmatrix}
\sim
\mathcal{N}_3
\begin{bmatrix}
0 \\
\vartheta \\
-\vartheta
\end{bmatrix}
\begin{bmatrix}
A_{11} & A_{12} & A_{13} \\
A_{21} & A_{22} & A_{23} \\
A_{31} & A_{32} & A_{33}
\end{bmatrix}.
$$
where \( \vartheta = -\sum R^\top (R \Sigma R^\top)^{-1} c \) with \( \Sigma = V_{\gamma\gamma}^{-1} \) which is the covariance matrix of \( \hat{\gamma} \).

Also \( A_{11} = \Sigma, A_{12} = A_{21}^\top = \Sigma - \Sigma (R \Sigma R^\top)^{-1} R \Sigma, A_{13} = A_{31}^\top = \Sigma R^\top (R \Sigma R^\top)^{-1} R \Sigma, A_{22} = A_{12}, A_{23} = A_{32}^\top = 0 \) and \( A_{33} = A_{13} \).

**Proof:** See Appendix.

First, we begin with the definition of the asymptotic distributional bias (ADB) of the estimators. For any estimator \( \gamma^* \), the ADB is defined as:

\[
ADB(\gamma^*) = E \left[ \lim_{n \to \infty} \sqrt{n}(\gamma^* - \gamma) \right].
\]  

(4.10)

We use some notations to succinct the ADB and ADR formulas. Let \( Z_1 = \chi^2_{l_2+2}(\Delta) \) and \( Z_2 = \chi^2_{l_2+4}(\Delta) \) be the non-central \( \chi^2 \) random variables. The distribution function of a non-central \( \chi^2 \) random variable with non-centrality parameter \( \Delta \) and degrees of freedom \( \nu \) is denoted by \( \Psi_{\nu}(x, \Delta) = \Pr(\chi^2_{\nu}(\Delta) \leq x) \). Assume \( \chi^2_{l_2,\alpha} \) is the level \( \alpha \) critical value of the central \( \chi^2 \) distribution with \( l_2 \) degrees of freedom.

**Theorem 3:** Using the above definition of ADB and Theorem 2, under the local alternatives \( K_{(n)} \) in (4.9) and the usual regularity conditions as \( n \to \infty \),

\[
ADB(\hat{\gamma}) = \vartheta
\]

\[
ADB(\hat{\gamma}_P) = \vartheta \Psi_{\nu+4}(\chi^2_{\nu+2,\alpha}, \Delta)
\]

\[
ADB(\hat{\gamma}_S) = \nu \vartheta E(Z_1^{-1})
\]

\[
ADB(\hat{\gamma}_{S+}) = ADB(\hat{\gamma}_S) + \vartheta \Psi_{\nu+4}(\nu, \Delta) - \vartheta \nu E \left( Z_1^{-1} I(Z_1 < \nu) \right)
\]
where \( \vartheta = -\Sigma R^T (R\Sigma R^T)^{-1} c \), \( Z_1 = \chi_{l_2, 2, \alpha}^2(\Delta) \), \( \nu = l_2 - 2 \), \( \Delta = c^T (R\Sigma R^T)^{-1} c \), and \( \Psi_g(x, \Delta) \) is the distribution function of a non-central chi-square random variable with \( g \) degrees of freedom and non-centrality parameter \( \Delta \).

**Proof:** See Appendix.

We now compare the ADBs of the four estimators. The ADBs of all estimators depend on \( \vartheta \) in terms of \( c \), as well as \( \Delta = 0 \). So it is sufficient to compare using \( \Delta \) only.

The ADB of \( \hat{\gamma} \) is an unbounded function of \( c \) as well as \( \Delta = c^T (R\Sigma R^T)^{-1} c \). On the other hand, the ADBs of \( \hat{\gamma}_P \), \( \hat{\gamma}_S \), and \( \hat{\gamma}_S^+ \) are bounded in \( \Delta \) as \( \text{E}(Z_1) \) is a decreasing function of \( \Delta \). The characteristics of \( \hat{\gamma}_P \) and \( \hat{\gamma}_S^+ \) are similar to that of \( \hat{\gamma}_S \).

As we calculate the quadratic bias (QB) numerically in the simulation study, we need the following formula to obtain a measurable analysis of the bias functions of any estimator \( \hat{\gamma}^* \):

\[
QB(\hat{\gamma}^*) = [\text{ADB}(\hat{\gamma}^*)]^T \Sigma [\text{ADB}(\hat{\gamma}^*)].
\]

Second, we define the ADRs of the estimators. To derive expressions for the ADRs of estimators, we define a quadratic loss function

\[
L(\gamma^*; M) = (\sqrt{n}(\gamma^* - \gamma))^T M (\sqrt{n}(\gamma^* - \gamma)),
\]

where \( M \) is a positive semi-definite weight matrix. In the model selection, we often use generic measure of model fitting but we are not interested in selecting of the weight matrix \( M \). In this case, we recommend \( l \times l \) identity matrix \( M = I_l \) as this
provides the quadratic loss. Many researchers use this identity matrix for their real data applications. We may choose other weight matrices but they do not guarantee that the shrinkage estimators perform better than the UMPLE estimator.

The expected loss function, or the risk function, is defined as:

$$ ADR(\gamma^*, \gamma; M) = \lim_{n \to \infty} E(\mathcal{L}(\gamma^*; M)). $$

The risk function can be written as:

$$ ADR(\gamma^*, \gamma; M) = \lim_{n \to \infty} E\left( (\sqrt{n}(\gamma^* - \gamma))^\top M (\sqrt{n}(\gamma^* - \gamma)) \right) = \text{tr} \left[ M \Sigma^*(\gamma^*) \right], \quad (4.11) $$

where $\Sigma^*(\gamma^*)$ is the asymptotic covariance matrix of $\gamma^*$.

Under the sequence of local alternatives, we define the asymptotic distribution function of an estimator $\gamma^*$ as:

$$ G(z) = \lim_{n \to \infty} P(\sqrt{n}(\gamma^* - \gamma) \leq z | K(n)), $$

where $G(z)$ is a nondegenerate distribution function for the estimators we consider.

We define the asymptotic distributional quadratic risk (ADR) by:

$$ ADR(\gamma^*; M) = \int \cdots \int z^\top M z dG(z) = \text{tr} \left[ M \Sigma^*(\gamma^*) \right], $$

where $\Sigma^*(\gamma^*) = \int \cdots \int zz^\top dG(z)$ is the dispersion matrix for the distribution $G(z)$. 
An estimator $\gamma^*$ dominates an estimator $\gamma^{**}$ asymptotically if $\text{ADR}(\gamma^*; M) \leq \text{ADR}(\gamma^{**}; M)$. If, in addition, $\text{ADR}(\gamma^*; M) < \text{ADR}(\gamma^{**}; M)$ for at least some $(\gamma, M)$, then $\gamma^*$ strictly dominates $\gamma^{**}$. This leads us to the following Theorem.

**Theorem 4:** Under the local alternatives $K_n$ in (4.9) and the usual regularity conditions, the ADR of the positive shrinkage estimator is:

$$
\text{ADR}(\hat{\gamma}_S^+; M) = \text{ADR}(\hat{\gamma}_S; M) - \text{trace}(A_{13}M) \mathbb{E}\left[\left(1 - \nu Z_1^{-1}\right)^2 I(Z_1 < \nu)\right] + \vartheta^\top M \vartheta \left(2 \Psi_{\nu+4}(\nu, \Delta) - 2\nu \mathbb{E}(Z_1^{-1} I(Z_1 < \nu))\right) - \vartheta^\top M \vartheta \mathbb{E}\left(\left(1 - \nu Z_2^{-1}\right)^2 I(Z_2 < \nu)\right),
$$

where

$$
\vartheta = -\Sigma R^\top (R \Sigma R^\top)^{-1} c, \quad A_{12} = \Sigma - \Sigma (R \Sigma R^\top)^{-1} R \Sigma, \quad A_{13} = \Sigma R^\top (R \Sigma R^\top)^{-1} R \Sigma,
$$

and $\nu = l_2 - 2$. 


Proof: See Appendix.

Under $H_0: \gamma_2 = 0$ or $\Delta = 0$, the gain in risk of all the estimators over the UMPLE $\hat{\gamma}$ is substantial, but the RMPLE $\tilde{\gamma}$ is the best choice in terms of lower ADR. If $\Delta > 0$, $\hat{\gamma}$ is not a good choice, as the ADB and ADR increase and ADR becomes unbounded. Thus, the performance of $\tilde{\gamma}$ depends strongly on the reliability of $H_0$. Under $H_0$, the difference of ADRs for $\hat{\gamma}$ and $\hat{\gamma}_P$ is $\text{trace}(M A_{13}) \Psi_{\nu+2}(\nu, 0)$. When $\Delta$ deviates from 0, the risk of $\hat{\gamma}_P$ monotonically approaches to the ADR of $\hat{\gamma}$ after achieving a maximum value. The ADR of $\hat{\gamma}_S$ is smaller or equal to the ADR of $\hat{\gamma}$ in the entire parameter space, and the upper limit is attained when $\Delta \to \infty$. The risk of $\hat{\gamma}_{S+}$ is lower than that of $\hat{\gamma}_S$ when $\Delta > 0$. Therefore, the positive-part shrinkage estimator dominates the shrinkage estimator. Hence, the positive-part shrinkage estimator performs better than the UMPLE.

In an effort to numerically evaluate the performance of proposed estimators, we conduct a simulation study in the next section to compare the performance of estimators for selected sample sizes.

4.5 Simulation results

To compare the performance of the restricted, pretest, and shrinkage estimators described in Sections 4.3 and 4.4, we have designed a Monte Carlo simulation study.
We illustrate the pretest and shrinkage methods by simulation for the special case of a binomial GARFIMA model with the logit link function. Thus, we present the simulation results based on the binary response model:

\[ \mu_t = \frac{\exp(\eta_t)}{1 + \exp(\eta_t)}. \]  

(4.12)

where \( \eta_t = x_t^\top \gamma + \tau_t \) in which \( \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_l)^\top \) is a set of regression parameters, and \( \tau_t = \sum_{j=1}^{p} \phi_j (g(y_{t-j}) - x_{t-j-1}^\top \gamma) + \sum_{k=0}^{\infty} \rho_k \rho_{t-k} \) is an ARFIMA process. In the simulation experiments, we generate \( \{x_t\}_{t=1}^{n} \) from the model specified by (4.3) and \( \varepsilon \sim N(0,1) \). We simulate 1000 replicates of a GARFIMA with autoregressive parameter \( \phi = 0.58 \), the amount of fractional differencing is \( d = 0.4 \), a moving average parameter \( \delta = 0.9 \), and samples sizes are \( n = 250 \) and \( n = 350 \). The ACF of \( x_t \) is given in Figure 4.1. We apply the logit as a link function, and 11 to 21 covariates are included in the simulations. An important aspect to assess in a time series is its stationarity, and we consider the value of \( d = 0.4 \) to make the ARFIMA process stationary.

![Figure 4.1: The ACF plot based on simulated data](image-url)
We consider a special case of the hypothesis $H_0 : R\gamma = 0$, where we partition the regression parameter vector into two sub-vectors $\gamma = (\gamma_1^T, \gamma_2^T)^T$, where $\gamma_1$ and $\gamma_2$ are assumed to have dimensions $l_1 \times 1$ and $l_2 \times 1$, respectively, such that $l = l_1 + l_2$. Our aim is to estimate $\gamma_1$ after incorporating the auxiliary information $\gamma_2 = 0$. True values of $\gamma = (\gamma_1^T, \gamma_2^T)^T$ are set to $(1.56, -0.80, 1.10, -0.91, -0.95)$, the intercept term is set to $\gamma_0 = 0.50$ and $\gamma_2 = 0$. We define the parameter, $\Delta = \| \gamma - \gamma^{(0)} \|_2$, to simulate binary response where $\| . \|$ is the Euclidean norm, $\Delta$ is a non-negative number, and $\gamma^{(0)} = (\gamma_1^T, 0^T)^T$. We assume that $\Delta = (0, 0.29, 0.57, 0.86, 1.14, 1.43, 1.71, 2.0)$ We provide detailed results for six cases: $(l_1, l_2) = (3, 8), (3, 12), (3, 18), (5, 6), (5, 10), \text{and} (5, 16)$. The simulation results are presented in Tables 4.1-4.2 and Figure 4.2, when the active set of parameters are $\gamma_1$ and the inactive set of parameters is $\gamma_2$.

The objective here is to investigate the behaviour of the proposed estimators for $H_0 : \Delta = 0$ and $\Delta \in (0, 2]$. The criterion for comparing the performance of any estimator $\hat{\gamma}_1^*$ in this study is the mean squared error (MSE). The simulated relative mean squared error (RMSE) of $\gamma_1^*$ to $\hat{\gamma}_1$ is defined as $\text{RMSE}(\hat{\gamma}_1 : \gamma_1^*) = \text{MSE}(\hat{\gamma}_1) / \text{MSE}(\gamma_1^*)$. Observe that an RMSE $> 1$ indicates the degree of superiority of $\gamma_1^*$ over $\hat{\gamma}_1$. 
\begin{table}
\centering
\begin{tabular}{ccccccc}
\hline
 & \(l_2 = 8\) & \(l_2 = 12\) & \(l_2 = 18\) & \(l_2 = 6\) & \(l_2 = 10\) & \(l_2 = 16\) \\
\hline
RMPLE & 1.49 & 1.66 & 2.15 & 1.46 & 1.64 & 2.10 \\
PT & 1.37 & 1.50 & 1.65 & 1.27 & 1.41 & 1.60 \\
SE & 1.32 & 1.47 & 1.77 & 1.21 & 1.37 & 1.67 \\
PSE & 1.34 & 1.49 & 1.81 & 1.24 & 1.47 & 1.76 \\
\hline
\end{tabular}
\caption{RMSEs of RMPLE, PT, SE, and PSE with respect to \(\hat{\gamma}_1\) when the restricted parameter space is correct (\(\Delta = 0\)), and \(n = 250\).}
\end{table}

\begin{table}
\centering
\begin{tabular}{ccccccc}
\hline
 & \(l_2 = 8\) & \(l_2 = 12\) & \(l_2 = 18\) & \(l_2 = 6\) & \(l_2 = 10\) & \(l_2 = 16\) \\
\hline
RMPLE & 1.39 & 1.51 & 1.77 & 1.35 & 1.48 & 1.71 \\
PT & 1.30 & 1.41 & 1.51 & 1.20 & 1.29 & 1.44 \\
SE & 1.26 & 1.39 & 1.57 & 1.14 & 1.28 & 1.47 \\
PSE & 1.28 & 1.43 & 1.60 & 1.18 & 1.35 & 1.50 \\
\hline
\end{tabular}
\caption{RMSEs of RMPLE, PT, SE, and PSE with respect to \(\hat{\gamma}_1\) when the restricted parameter space is correct (\(\Delta = 0\)), and \(n = 350\).}
\end{table}
We summarize our findings from Tables 4.1-4.2, and Figure 4.2

(i) The numerical values of RMSE are summarized in Tables 4.1-4.2 and Figure 4.2. The RMPLE outperforms all the other estimators at and near $\Delta = 0$, i.e., $H_0 : \gamma_2 = 0$ subject to some sampling fluctuation in the data generating. On the contrary, when $\Delta$ increases, the estimated RMSE of $\hat{\gamma}_1$ decreases. That is, the risk of $\hat{\gamma}_1$ becomes unbounded, whereas the estimated RMSEs of all other estimators converge to 1. It can be safely concluded that the digressing from $H_0$.
is critical to RMPLE, $\tilde{\gamma}_1$. This corroborates our theoretical risk of the estimators in Theorem 4, as the value of RMSE less than 1 means the asymptotic risk ADR of RMPLE is higher than the reference estimator UMLE. The pretest estimator performs better at and near $H_0$, but the simulation shows that the performance heavily depends on how close $\Delta$ is to zero.

(ii) Both SE and PSE estimators have lower risk than UMLE That is, the RMSE values of SE and PSE are greater than 1. The PSE outperforms SE for every combination of $n$, $l_2$, and $\Delta$ except for some sampling fluctuations. The severity of risk (or ADR) of SE and PSE mainly depends on the number of inactive covariates $l_2$ and the value of $\Delta$. Figure 3.2 show that the risk behaviour of the estimators for large values of $l_2$ and $\Delta(0 \leq \Delta \leq 2)$.

(iii) When $l_1 = 3, 5$ and $l_2$ varies from 6 to 18, the RMSEs are higher for $l_1 = 3$ as $l_2$ varies compared to when $l_1 = 5$ as $l_2$ varies. It proves that the RMSEs of SE and PSE with respect to UMLE increase for increasing $l_2$. For example, if $\Delta = 0$, $l_1 = 3, l_2 = 18$ and $n = 250$, then the RMSEs of SE and PSE are 1.77 and 1.81, respectively. On the other hand, if $\Delta = 0$, $l_1 = 5, l_2 = 16$ and $n = 250$, then the RMSEs of SE and PSE are 1.67 and 1.76, respectively. That is, the shrinkage method produces improved estimates when $l_2$ is large and $\Delta$ is near zero.

In summary, a) the PSE have smaller MSE than the PT and SE for moderate and
large number inactive covariates and b) there is a substantial difference between the UMPLE and the RMLE at and near $\Delta = 0$.

### 4.6 Real Data Application

To exemplify the usefulness of the GARFIMA, we illustrate the pretest and shrinkage estimation strategies using a financial time series data with a binary response. This dataset obtained from http://finance.yahoo.comYahoo Finance cover the monthly closing price of S&P500 stock market index from the beginning of the 1970s to the end of 2019. The monthly returns can be computed as $MR_t = 100 \times [\log(P_t) - \log(P_{t-1})]$, where $P_t$ denotes the monthly closing index price. Now the monthly excess returns on the S&P500 index are computed by subtracting the 3-month U.S. Treasury Bill rate ($TB_t$) rate from the monthly returns on the S&P500 as $r_t = MR_t - TB_t$, $t = 1, 2, \cdots, 600$. The treasury bill rate data are obtained from https://fred.stlouisfed.org/series/TB3MS. In directional prediction, the binary response variable is created from the monthly excess return series as $y_t = 1$, if the sign of investment growth (or monthly excess returns) is positive and 0, otherwise. An increasing amount of empirical evidence suggests out-of-sample predictability of excess monthly stock market returns can be predicted by a range of financial and macroeconomic variables. Christoffersen (2006)[11] claimed that the sign of stock returns could be predictable as much as returns are not themselves predictable. In this
connection, Nyberg (2010)[47] implemented a dynamic autoregressive binary probit regression model using the sign of returns on the S&P500 and found that his model performs better than a number of other models based on a continuous response in terms of out-of-sample sign forecasts. Our model (4.12) with ARFIMA(1, d, 1) in the linear predictor $\eta_t$ may offer a parsimonious way to take the longer history of lagged values of $P_t$ and TB$_t$ which are important covariates in explaining the monthly excess return as either positive or negative. We are interested in examining the predictive power of excess monthly stock market returns based on many lagged values of monthly closing index price and treasury bill rate.

Figure 4.3: Time series plots of (a) monthly excess stock return, (b) the S&P500 closing price, and (c) the 3-month U.S. Treasury Bill
Hossain (2016)[32] and Leung (2000)[37] considered the model \( \eta_t = \mathbf{x}_t^\top \gamma, \quad t = 1, \cdots, n \) for time series data. We consider the GARFIMA with eight lagged values of \( P_t \) and \( TB_t \) for binary response \( r_t \). Let \( \mathbf{x}_{t-1} = (1, TB_{t-1}, TB_{t-2}, \cdots, TB_{t-8}, P_{t-1}, P_{t-2}, \cdots, P_{t-8})^\top \) be the vector of covariates, and let \( \gamma = (\gamma_0, \gamma_1, \gamma_2, \cdots, \gamma_{16})^\top \) be the vector of regression parameters. Due to the use of eight lagged responses as covariates, we are constrained to take \( t = 9, 10, \cdots, 600 \), so the total number of observations in the analysis becomes \( n = 592 \).

Figure 4.3 provides a time series plot of \( r_t, P_t, \) and \( TB_t \). The plot (a) reveals a very distinctive seasonality which we shall incorporate into the ARFIMA model. We begin our analysis by examining the data. The raw time series of the monthly excess return is shown in Figure 4.3(a) which reveals a very distinctive seasonality that we then incorporate into the GARFIMA model. Observe the extremely high return at index 380 in Figure 4.3(a). Observe that the series \( r_t \) is frequently negative and the proportion of zeroes is 84%. Figures 4.3(b) and 4.3(c) are plots of the raw data, which exhibit downward trends.

To implement the pretest and shrinkage methods, we need to find the restricted model. The stepwise method with the AIC criteria suggests that the covariates \( TB_{t-1}, P_{t-4}, P_{t-5}, P_{t-6}, \) and \( P_{t-7} \) are the active covariates for predicting monthly excess return, and that the auxiliary information of remaining \( l_2 = 11 \) covariates will be used to estimate the coefficients of 5 active covariates (\( l_1 = 5 \)). Recall from the
simulation study, we obtain the inactive set of coefficients \( H_0 : \gamma_2 = 0 \), where \( \gamma_2 = (\gamma_6, \gamma_7, \cdots, \gamma_{16})^\top \) is a \( 11 \times 1 \) vector.

The sample autocorrelation function (ACF) and partial ACF (PACF) are useful qualitative tools to assess the presence of autocorrelation at individual lags. Figure 4.4 show the ACF and PACF plots of the standardized Pearson residuals for the RMPLE model. These plots suggest that the residuals behave as white noise.

![Figure 4.4: (a) Autocorrelation function (ACF) plot of RMPLE model model. (b) Partial autocorrelation function (PACF) plot of the RMPLE model for the S&P500 data.](image)

As we have one dataset, we use case the resampling bootstrap method to obtain point estimates, standard errors, and RMSEs. We apply the following bootstrap algorithm:

i. Set the bootstrap sample of size \( n = 500 \).
ii. Fit the GARFIMA according equation (4.12) and obtain the UMPLE \(\hat{\gamma}_1\).

iii. Calculate the fitted values of the model, \(\hat{\mu}_t\).

iv. Generate bootstrap replicates \(\{y_t^*\}_{t=1}^n\) with mean \(\hat{\mu}_t\).

v. Estimate \(\gamma_1\) based on \(\{y_t^*\}_{t=1}^n\) samples.

We obtained 5000 bootstrap samples using this algorithm and calculated the estimates, standard errors, and RMSEs. The results given in Table 4.3 reveal that the restricted, pretest, shrinkage, and positive shrinkage estimators are superior to the unrestricted maximum partial likelihood estimator, which is in agreement with our simulation results.
Table 4.3: Estimates (first row) and standard errors (second row) for $T_{t-1}(\tilde{\gamma}_1)$, $P_{t-4}(\tilde{\gamma}_2)$, $P_{t-5}(\tilde{\gamma}_3)$, $P_{t-6}(\tilde{\gamma}_4)$, $P_{t-7}(\tilde{\gamma}_5)$

<table>
<thead>
<tr>
<th>Method</th>
<th>Estimate</th>
<th>SE</th>
<th>SE</th>
<th>SE</th>
<th>SE</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>UMLE</td>
<td>-0.732</td>
<td>0.013</td>
<td>0.023</td>
<td>-0.021</td>
<td>0.015</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>(0.866)</td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.012)</td>
<td>(0.011)</td>
<td></td>
</tr>
<tr>
<td>RMLE</td>
<td>-0.380</td>
<td>0.012</td>
<td>0.016</td>
<td>-0.018</td>
<td>0.013</td>
<td>2.342</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.007)</td>
<td>(0.009)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td></td>
</tr>
<tr>
<td>PT</td>
<td>-0.675</td>
<td>0.013</td>
<td>0.021</td>
<td>-0.020</td>
<td>0.014</td>
<td>1.004</td>
</tr>
<tr>
<td></td>
<td>(0.740)</td>
<td>(0.009)</td>
<td>(0.011)</td>
<td>(0.011)</td>
<td>(0.010)</td>
<td></td>
</tr>
<tr>
<td>SE</td>
<td>-0.613</td>
<td>0.013</td>
<td>0.020</td>
<td>-0.020</td>
<td>0.014</td>
<td>1.913</td>
</tr>
<tr>
<td></td>
<td>(0.537)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td></td>
</tr>
<tr>
<td>PSE</td>
<td>-0.611</td>
<td>0.013</td>
<td>0.020</td>
<td>-0.020</td>
<td>0.014</td>
<td>1.928</td>
</tr>
<tr>
<td></td>
<td>(0.536)</td>
<td>(0.008)</td>
<td>(0.010)</td>
<td>(0.010)</td>
<td>(0.009)</td>
<td></td>
</tr>
</tbody>
</table>

4.7 Conclusion

In this paper, we proposed the pretest and shrinkage estimators for GARFIMA models, when it is suspected that some of the coefficients of inactive covariates may be restricted to a subspace. We provided the ADB and ADR expressions of the pretest and shrinkage estimators. We conducted Monte Carlo simulations to examine the performance of the estimators which showed that the RMLE outperforms the usual UMLE at or near the restriction. However, as we deviate from the restriction,
the RMSE decreases; that is, the risk of RMLE increases. Moreover near the restriction, the risk of the pretest estimator is lower than that of the UMLE. In fact, the shrinkage estimators outperform the UMLE in the entire parameter space when the number of inactive parameter is greater than two. In addition, the performance of the shrinkage estimators improve relative to the UMLE when the number of inactive covariates increases.

Finally, we applied the proposed strategies to S&P500 data to evaluate the relative performance of the proposed estimators. The results are consistent with our analytical and simulated findings.

In terms of recommendations, it shows that simulation and real data results justify the better performance of shrinkage estimators, in terms of higher accuracy and lower variability, in the estimation of regression parameters for generalized ARFIMA model. If we are uncertain of the quality of auxiliary information about the covariates, we can still use the shrinkage estimation because it offers a lower or at best an equal risk relative to the UMLE over the entire parameter space.

4.8 Appendix

Proof of Theorem 1

The proof follows the same lines as the proof of Theorem 3.1 in Fokianos (1998)[20].
The key point is to show that the partial score vector

\[ S_t(\beta) = \frac{\partial \ell(\beta)}{\partial \beta} = \sum_{s=1}^{t} x_{s-1} \cdot \frac{\partial \mu_s}{\partial \eta_s} \cdot \frac{y_s - \mu_s(\beta)}{\sigma_s^2(\beta)} \cdot \frac{\partial ^2 \ell(\beta)}{\partial \beta \partial \beta^\top}, \quad t = 1, 2, \cdots, n \]

is a zero mean square integrable martingale sequence with respect to an adequate filtration \( \{\mathcal{F}_t\}; t = 1, 2, \cdots, n \). It is straightforward to show that, under the regularity conditions (1) and (2), \( E(S_{t+1}(\beta)|\mathcal{F}_t) = S_t(\beta) \) as \( E(y_{t+1} - \mu_{t+1}(\beta)|\mathcal{F}_t) = 0 \). Using this fact, we can show that the asymptotic distribution of the partial score converges to a normal random variable \( \frac{S_n(\beta)}{\sqrt{n}} \rightarrow N_{l+p+q+1}(0, G(\beta)) \) in distribution as \( n \rightarrow \infty \), where \( \frac{G_n(\beta)}{n} \rightarrow G(\beta) \) in probability and \( G(\beta) \) is a positive definite matrix at the true values \( \beta \) and therefore its inverse exists. The proof of this fact is based on verifying the conditions of corollary 3.1 in Hall (1980)[28].

The asymptotic properties of the MPLE is to show that the remainder term \( H_n(\beta) = G_n(\beta) - R_n(\beta) \) is small where \( H_n(\beta) = -\frac{\partial^2 \ell(\beta)}{\partial \beta \partial \beta^\top} \). It can be shown that \( \frac{R_n(\beta)}{\sqrt{n}} \rightarrow 0 \) in probability as \( n \rightarrow \infty \), see Fokianos (1998)[20].

We observe that a Taylor series expansion of \( S_n(\beta) = 0 \) around \( \hat{\beta} \) yields the following useful approximation up to terms asymptotically negligible in probability,

\[ \sqrt{n}(\hat{\beta} - \beta) \approx \left( \frac{H_n(\beta)}{n} \right)^{-1} S_n(\beta) \] but \( \frac{H_n(\beta)}{n} = \frac{G_n(\beta) - R_n(\beta)}{n} = \frac{G_n(\beta)}{n} = G(\beta) \) as \( n \rightarrow \infty \).

Thus we get \( \sqrt{n}(\hat{\beta} - \beta) \approx G^{-1}(\beta) \frac{S_n(\beta)}{\sqrt{n}} \), which together with an appeal to Slutsky’s
Theorem leads to the required results of Theorem 1.

**Proof of Theorem 2**

\[ E(\sqrt{n}(\hat{\gamma} - \gamma)) = 0. \]

\[ E(\sqrt{n}(\hat{\gamma} - \gamma)) = E(\sqrt{n}(\hat{\gamma}_1 - \gamma_1 - \Sigma R^T (R\Sigma R^T)^{-1}\hat{\gamma}_2)) = -\Sigma R^T (R\Sigma R^T)^{-1}\sqrt{n}(e/\sqrt{n}) = \theta. \]

\[ E(\sqrt{n}(\hat{\gamma} - \tilde{\gamma})) = E(\sqrt{n}(\hat{\gamma} - \gamma) - \sqrt{n}(\tilde{\gamma} - \gamma)) = -\theta. \]

\[ \text{Var}(\sqrt{n}(\hat{\gamma} - \gamma)) = \Sigma - \Sigma(R\Sigma R^T)^{-1}R\Sigma = A_{11}. \]

\[ \text{Var}(\sqrt{n}(\hat{\gamma} - \gamma)) = \text{Var}(\sqrt{n}(\hat{\gamma}_1 - \gamma_1 - \Sigma R^T (R\Sigma R^T)^{-1}\hat{\gamma}_2)) = \Sigma - \Sigma(R\Sigma R^T)^{-1}R\Sigma = A_{22}. \]

\[ \text{Var}(\sqrt{n}(\hat{\gamma} - \tilde{\gamma})) = \text{Var}(\sqrt{n}(\hat{\gamma} - \gamma) - \sqrt{n}(\tilde{\gamma} - \gamma)) = \Sigma R^T (R\Sigma R^T)^{-1}R\Sigma = A_{33}. \]

\[ \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\tilde{\gamma} - \gamma)) = \Sigma - \Sigma(R\Sigma R^T)^{-1}R\Sigma = A_{12} = A_{21}^T. \]

\[ \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\tilde{\gamma} - \gamma)) \]

\[ = \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\hat{\gamma} - \gamma) - \sqrt{n}(\tilde{\gamma} - \gamma)) = \Sigma R^T (R\Sigma R^T)^{-1}R\Sigma = A_{13} = A_{31}^T. \]

\[ \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\tilde{\gamma} - \gamma)) = \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\gamma - \gamma) - \sqrt{n}(\tilde{\gamma} - \gamma)) \]

\[ = \text{Cov}(\sqrt{n}(\hat{\gamma} - \gamma), \sqrt{n}(\gamma - \gamma)) - \text{Cov}(\sqrt{n}(\tilde{\gamma} - \gamma), \sqrt{n}(\gamma - \gamma)) = 0. \]

**Proof of Theorem 3**

In this proof, we derive the bias expressions of the proposed estimators. It is obvious
that $\text{ADB}(\hat{\gamma}) = 0$. The ADB of RMPLE ($\hat{\gamma}$), PT ($\hat{\gamma}_P$), SE ($\hat{\gamma}_S$), and PSE ($\hat{\gamma}_{S+}$) estimators are as follows:

$$\text{ADB}(\hat{\gamma}) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \gamma)\right) = -\Sigma R^T (R \Sigma R^T)^{-1} c = \vartheta$$

$$\text{ADB}(\hat{\gamma}_P) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_P - \gamma)\right) = E\left(\lim_{n \to \infty} (\sqrt{n} I \left(\hat{D}_n \leq \chi^2_{l_2,\alpha}\right) (\hat{\gamma} - \gamma)\right) = \vartheta \Psi_{\nu+1}(\chi^2_{\nu+2,\alpha}; \Delta)$$

$$\text{ADB}(\hat{\gamma}_S) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_S - \gamma)\right) = -E\left(\lim_{n \to \infty} \sqrt{n}(l_2 - 2) \hat{D}_n^{-1}(\hat{\gamma} - \gamma)\right) = \nu \vartheta E(Z^{-1})$$

$$\text{ADB}(\hat{\gamma}_{S+}) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_{S+} - \gamma)\right)$$

$$= E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_S - \gamma) - \sqrt{n} \left(1 - (l_2 - 2) \hat{D}_n^{-1}\right) I \left(\hat{D}_n < (l_2 - 2)\right) (\hat{\gamma} - \gamma)\right)$$

$$= \text{ADB}(\hat{\gamma}_S) - \vartheta E\left(I \left(\hat{D}_n < (l_2 - 2)\right) \left(1 - (l_2 - 2) \hat{D}_n^{-1}\right)\right)$$

$$= \text{ADB}(\hat{\gamma}_S) + \vartheta \Psi_{\nu+1}(\nu, \Delta) - \vartheta \nu E(Z^{-1} I(Z_1 < \nu))$$

**Proof of Theorem 4**

Based on the definition of ADR function, it is necessary to derive the asymptotic covariance matrices for the four estimators. The covariance matrix of any estimator $\hat{\gamma}^*$ is defined as:

$$\text{Cov}(\hat{\gamma}^*) = E\left(\lim_{n \to \infty} n(\hat{\gamma}^* - \gamma)(\hat{\gamma}^* - \gamma)^T\right).$$

First, we will start deriving the covariance matrices of the URMPLE and RMPLE:

$$\text{Cov}(\hat{\gamma}) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T\right) = \mathbf{A}_{11}$$

$$\text{Cov}(\hat{\gamma}) = E\left(\lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T\right) = \mathbf{A}_{12} + \vartheta \vartheta^T.$$
Second, we derive the covariance matrix of the pretest estimator:

\[
\text{Cov}(\hat{\gamma}_P) = \mathbb{E} \left( \lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_P - \gamma) \sqrt{n}(\hat{\gamma}_P - \gamma)^\top \right)
\]

\[
= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma) \sqrt{n}(\hat{\gamma} - \gamma)^\top + \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2, \alpha} \right) \right)
\]

\[
- 2\mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2, \alpha} \right) \right)
\]

\[
= \mathcal{A}_{11} + \mathcal{A}_{13} \Psi_{\nu+4} \left( \chi^2_{\nu+2, \alpha}, \Delta \right) + \Psi^\top \Psi_{\nu+6} \left( \chi^2_{\nu+2, \alpha}, \Delta \right)
\]

\[
- 2\mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2, \alpha} \right) \right).
\]
Consider the fourth term:

\[
\begin{align*}
&\mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{E} \left( \mathbb{E} \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) | \sqrt{n}(\hat{\gamma} - \gamma) \right) \\
&= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
&= \mathbb{A}_{13} \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) + \vartheta \vartheta^\top \Psi_{\nu+6} (\chi^2_{\nu+2,\alpha}, \Delta) - \vartheta \vartheta^\top \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) \\
\end{align*}
\]

\[
\text{Cov}(\hat{\gamma}_p) = \mathbb{A}_{11} + \lim_{n \to \infty} \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
- 2 \lim_{n \to \infty} \mathbb{E} \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \lim_{n \to \infty} I \left( \hat{D}_n \leq \chi^2_{l_2,\alpha} \right) \right) \\
= \mathbb{A}_{11} + \mathbb{A}_{13} \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) + \vartheta \vartheta^\top \Psi_{\nu+6} (\chi^2_{\nu+2,\alpha}, \Delta) \\
- 2 \left( \mathbb{A}_{13} \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) + \vartheta \vartheta^\top \Psi_{\nu+6} (\chi^2_{\nu+2,\alpha}, \Delta) \\
- \vartheta \vartheta^\top \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) \right) \\
= \mathbb{A}_{11} - \mathbb{A}_{13} \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta) - \vartheta \vartheta^\top \Psi_{\nu+6} (\chi^2_{\nu+2,\alpha}, \Delta) \\
+ 2 \vartheta \vartheta^\top \Psi_{\nu+4} (\chi^2_{\nu+2,\alpha}, \Delta).
\]

Third, we derive the covariance matrices of the shrinkage and positive shrinkage.
estimators:

\[
\text{Cov}(\hat{\gamma}_S) = \mathbb{E}\left( \lim_{n \to \infty} \sqrt{n}(\hat{\gamma}_S - \gamma) \sqrt{n}(\hat{\gamma}_S - \gamma)^T \right)
\]

\[
= \mathbb{E}\left( \lim_{n \to \infty} \sqrt{n} \left( \hat{\gamma} - \gamma - \nu \hat{D}_n^{-1}(\hat{\gamma} - \hat{\gamma}) \right) \sqrt{n} \left( \hat{\gamma} - \gamma - \nu \hat{D}_n^{-1}(\hat{\gamma} - \hat{\gamma}) \right)^T \right)
\]

\[
= \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T) + \nu^2 \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-2})
\]

\[-2\nu \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-1})
\]

\[
= \mathbf{A}_{11} + \nu^2 \mathbf{A}_{13} \mathbb{E}(Z_1^{-2}) + \nu^2 \mathbf{\vartheta}^T \mathbf{\vartheta} \mathbb{E}(Z_2^{-2}) - 2\nu \mathbb{E}(\sqrt{n}(\hat{\gamma} - \hat{\gamma})\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-1})
\]

Consider the last term:

\[
\mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-1}) = \mathbb{E}\left( \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-1} | \sqrt{n}(\hat{\gamma} - \gamma) \right)
\]

\[
= \mathbb{E}\left( \sqrt{n}(\hat{\gamma} - \gamma) \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)^T | \sqrt{n}(\hat{\gamma} - \gamma)) \lim_{n \to \infty} \hat{D}_n^{-1} \right)
\]

\[
+ \mathbb{E}\left( \sqrt{n}(\hat{\gamma} - \gamma) (\sqrt{n}(\hat{\gamma} - \gamma)^T - \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)^T) \lim_{n \to \infty} \hat{D}_n^{-1} \right)
\]

\[
= \mathbb{E}\left( \sqrt{n}(\hat{\gamma} - \gamma)\sqrt{n}(\hat{\gamma} - \gamma)^T \lim_{n \to \infty} \hat{D}_n^{-1} \right) - \mathbb{E}\left( \sqrt{n}(\hat{\gamma} - \gamma) \lim_{n \to \infty} \hat{D}_n^{-1} \right) \mathbb{E}(\sqrt{n}(\hat{\gamma} - \gamma)^T)
\]

\[
= \mathbf{A}_{13} \mathbb{E}(\chi_{1+2}^2(\Delta)) + \mathbf{\vartheta}^T \mathbb{E}(\chi_{1+4}^2(\Delta)) - \mathbf{\vartheta}^T \mathbb{E}(\chi_{1+2}^{-2}(\Delta))
\]

\[
= \mathbf{A}_{13} \mathbb{E}(Z_1^{-1}) + \mathbf{\vartheta}^T \mathbb{E}(Z_2^{-1}) - \mathbf{\vartheta}^T \mathbb{E}(Z_1^{-1})
\]
Hence
\[
\text{Cov}(\hat{\gamma}_S) = \mathcal{A}_{11} + \nu^2 \left( \mathcal{A}_{13} E(Z_1^{-2}) + \vartheta \vartheta^\top E(Z_2^{-2}) \right) \\
-2\nu \left( \mathcal{A}_{13} E(Z_1^{-1}) + \vartheta \vartheta^\top E(Z_2^{-1}) - \vartheta \vartheta^\top E(Z_1^{-1}) \right) \\
= \mathcal{A}_{11} + \left( \nu^2 E(Z_1^{-2}) - 2\nu E(Z_1^{-1}) \right) \mathcal{A}_{13} \\
+ \left( \nu^2 E(Z_2^{-2}) + 2\nu E(Z_1^{-1}) - 2\nu E(Z_2^{-1}) \right) \vartheta \vartheta^\top
\]

Let \( F_m(\Delta) = \left(1 - \nu \hat{D}_n^{-1}\right)^m I \left( \hat{D}_n < \nu \right) \), where \( m = 1, 2 \)

\[
\text{Cov}(\hat{\gamma}_{S+}) = \mathbb{E} \left( \lim_{n \to \infty} \sqrt{n} (\hat{\gamma}_{S+} - \gamma) \sqrt{n} (\hat{\gamma}_{S+} - \gamma)^\top \right) \\
= \mathbb{E} \left( \lim_{n \to \infty} \sqrt{n} (\hat{\gamma}_S - \gamma) \sqrt{n} (\hat{\gamma}_S - \gamma)^\top \right) \\
+ \mathbb{E} \left( \lim_{n \to \infty} F_2(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) \sqrt{n} (\hat{\gamma} - \tilde{\gamma})^\top \right) \\
-2\mathbb{E} \left( \lim_{n \to \infty} F_1(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) \sqrt{n} (\tilde{\gamma} - \gamma)^\top \right) \\
= \text{Cov}(\hat{\gamma}_S) + \mathbb{E} \left( \lim_{n \to \infty} F_2(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) \sqrt{n} (\hat{\gamma} - \gamma)^\top \right) \\
-2\mathbb{E} \left( \lim_{n \to \infty} F_1(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) \left( \sqrt{n} (\tilde{\gamma} - \gamma)^\top + \left(1 - \nu \hat{D}_n^{-1}\right) \sqrt{n} (\hat{\gamma} - \gamma)^\top \right) \right),
\]

\[
= \text{Cov}(\hat{\gamma}_S) + \mathbb{E} \left( \lim_{n \to \infty} F_2(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) \sqrt{n} (\hat{\gamma} - \gamma)^\top \right) \\
-2\mathbb{E} \left( \lim_{n \to \infty} F_1(\Delta) \sqrt{n} (\hat{\gamma} - \tilde{\gamma}) (\sqrt{n} (\tilde{\gamma} - \gamma)^\top \right).
\]
Consider the second term:

\[-E \left( \lim_{n \to \infty} F_2(\Delta) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})\sqrt{n}(\hat{\gamma} - \tilde{\gamma})^\top \right)\]

\[= -E \left( \lim_{n \to \infty} \left( 1 - \nu \hat{D}_n^{-1} \right)^2 I(\hat{D}_n < \nu) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})\sqrt{n}(\hat{\gamma} - \tilde{\gamma})^\top \right)\]

\[= -\mathcal{A}_{13} E \left( I(Z_1 < \nu) \left( 1 - \nu Z_1^{-1} \right)^2 \right) - \vartheta \vartheta^\top E \left( I(Z_2 < \nu) \left( 1 - \nu Z_2^{-1} \right)^2 \right) \]

Consider the third term:

\[-2E \left( \lim_{n \to \infty} F_1(\Delta) \sqrt{n}(\hat{\gamma} - \tilde{\gamma})(\sqrt{n}(\hat{\gamma} - \gamma)^\top) \right)\]

\[= -2E \left( \lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) E \left( F_1(\Delta) \sqrt{n}(\hat{\gamma} - \gamma)^\top | \sqrt{n}(\hat{\gamma} - \gamma) \right) \right)\]

\[= -2E \left( \lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) E \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \right) F_1(\Delta) + 0 \right)\]

\[= -2E \left( \lim_{n \to \infty} \sqrt{n}(\hat{\gamma} - \tilde{\gamma}) I(\hat{D}_n < \nu) - \nu \hat{D}_n^{-1} \sqrt{n}(\hat{\gamma} - \gamma) I(\hat{D}_n < \nu) \right) \]

\[\times \ E \left( \sqrt{n}(\hat{\gamma} - \gamma)^\top \right)\]

\[= 2\Psi_{\nu+4}(\nu, \Delta) \vartheta \vartheta^\top - 2\nu E \left( Z_1^{-1} I(Z_1 < \nu) \right) \vartheta \vartheta^\top \]

\[= \left( 2\Psi_{\nu+4}(\nu, \Delta) - 2\nu E \left( Z_1^{-1} I(Z_1 < \nu) \right) \right) \vartheta \vartheta^\top \]

Finally,

\[
\text{Cov} (\hat{\gamma}_{S+}) = \text{Cov} (\hat{\gamma}_S) - E \left( (1 - \nu Z_1^{-1})^2 I(Z_1 < \nu) \right) \mathcal{A}_{13} \\
+ \left( 2\Psi_{\nu+4}(\nu, \Delta) - 2\nu E \left( Z_1^{-1} I(Z_1 < \nu) \right) \right) \vartheta \vartheta^\top \\
- E \left( (1 - \nu Z_2^{-1})^2 I(Z_2 < \nu) \right) \vartheta \vartheta^\top.
\]
The ADR expressions in Theorem 4 now follow from equation (4.11) which completes the proof.
Chapter 5

Conclusion and Discussion

5.1 Conclusion and Future Work

In this dissertation, we concentrate on the pretest and shrinkage estimation strategies for some time series regression models. We apply these strategies to improve the performance of unrestricted estimators when the NSI/auxiliary of the covariates is available. Estimators based on the pretest, improved pretest, and shrinkage rules are presented in the chapters of the thesis, which incorporate the auxiliary information, as well as the level of confidence in the NSI/auxiliary information. We will evaluate the comparative properties of the restricted, pretest, improved pretest, and shrinkage estimators with the unrestricted estimator.

The estimation of parameters is considered in various regression models under the following time series data situations:
To describe autocorrelated error structures in the time series data and fit the regression model with autoregressive integrated moving average time series errors.

- Short-range dependence high frequency time series data and consider the Generalised autoregressive moving average (GARMA) model when the data type is count and binary.

- Long memory high-frequency time series data and extend GARMA model to non-gaussian ARFIMA time series models.

For all of the above cases, the relative performance of the estimators (restricted, pretest, improved pretest, and shrinkage) is discussed in details theoretically as well as computationally. These three cases are discussed in Chapters 2, 3, and 4, respectively.

Each chapter derives the ADB and ADR of the proposed estimators, then comparative properties of the estimators are studied. The weighted quadratic loss function is used to calculate the ADR. The scope in the parameter space is evaluated in which the proposed estimators are superior to the unrestricted estimator. The relative mean squared error which is the inverse of risk is used as a criterion for comparing the restricted, pretest, shrinkage, positive shrinkage, and improved pretest estimators with respect to the unrestricted estimator. The results of proposed the estimators are provided theoretically, computationally and graphically. In the following setting, we summarize our findings.
In Chapter 2, we studied the restricted, shrinkage, and positive-part shrinkage estimators in the context of a linear regression model with ARIMA error. This work is an extension of Ghahramani (2016)[25]. The risk performance of the estimators is investigated through asymptotic distributional biases, risks, and Monte Carlo simulation studies. We found that the restricted estimator performs best when the restriction is correct. On the other hand, the shrinkage, and positive shrinkage estimators outperform the unrestricted estimator in the entire parameter space.

In Chapter 3, we have proposed shrinkage and improved pretest estimation methods for the regression coefficients GARMA models. We extended the results in Hossain (2016)[32] for shrinkage and improved pretest estimation methods of GARMA models. We have applied the GLM framework to fit the UMLE and RMLE of GARMA models. We also studied the asymptotic distributional biases and risks for the proposed estimators and implemented a dominance for improved pretest and shrinkage estimators to dominate the UMLE. We performed extensive simulation studies for binary and count time series data, together with a real data example of S&P500 data set. The results of both studies showed that the improved pretest and shrinkage estimators outperform UMLE and perform better than RMLE when the restriction is seriously violated.

Chapter 4 is an extension of Chapter 3 from short-range dependence to long-range dependence time series structure namely, generalized GARMFIMA models. We
proposed the pretest and shrinkage estimators for GARFIMA models, when it is suspected that some of the coefficients of lie on a subspace. We provided the ADB and ADR expressions of the pretest and shrinkage estimators and compare them analytically. We conducted Monte Carlo simulations to examine the performance of the estimators which showed that the RMPLE outperforms the usual UMPLE at or near the restriction. The performance of the pretest estimator depends on the validity of the restriction. The shrinkage estimator performs better than the UMPLE in the whole parameter space.

5.1.1 Future Scope

There are possibilities of extending our works in the following ways:

• In Chapter 2, we compared shrinkage and positive shrinkage estimators in a linear regression model with ARIMA error. This study can be extended for ridge-type shrinkage and ridge-type positive part-shrinkage estimators in regression models with ARIMA error.

• In Chapter 3, we considered shrinkage estimation for the GARMA Model. We anticipate that these estimation methods can extend other types of models, such as integer-autoregressive (INAR) and binomial-mixed Poisson INAR. Also, in this chapter, we applied shrinkage estimation strategy for low-dimensional ($p < n$), such as $100 \leq n \leq 350$ and $8 \leq p \leq 24$. We can also consider an
extension of the shrinkage estimation strategy for high-dimensional \((p < n)\) and ultra-high-dimensional \((p > n)\).

- For our future research in Chapter 4, we will consider comparing the performance of pretest, shrinkage, and positive-part shrinkage for GARFIMA-GARCH time series models. The results will compare with existing results for GARFIMA models. In addition, future research will explore the characteristics of the shrinkage estimation method to the GARFIMA-GARCH models.
Bibliography


with ARMA errors Non-penalty shrinkage estimation of random effect models for longitudinal data with AR(1) errors, 1–18.


