Clemson University
TigerPrints

All Dissertations | Dissertations

5-2022

Penalized Estimation of Autocorrelation

Xiyan Tan
xtan@clemson.edu

Follow this and additional works at: https://tigerprints.clemson.edu/all_dissertations

 Part of the Longitudinal Data Analysis and Time Series Commons

Recommended Citation

Tan, Xiyan, "Penalized Estimation of Autocorrelation" (2022). *All Dissertations*. 3016.
https://tigerprints.clemson.edu/all_dissertations/3016

This Dissertation is brought to you for free and open access by the Dissertations at TigerPrints. It has been accepted for inclusion in All Dissertations by an authorized administrator of TigerPrints. For more information, please contact kokeefe@clemson.edu.

Penalized Estimation of Autocorrelation

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Statistics

---

by
Xiyan Tan
May 2022

Accepted by:
Dr. Colin Gallagher, Committee Chair
Dr. Andrew Brown
Dr. Christopher McMahan
Dr. Qiong Zhang

Abstract

This dissertation explored the idea of penalized method in estimating the autocorrelation (ACF) and partial autocorrelation (PACF) in order to solve the problem that the sample (partial) autocorrelation underestimates the magnitude of (partial) autocorrelation in stationary time series. Although finite sample bias corrections can be found under specific assumed models, no general formulae are available. We introduce a novel penalized M-estimator for (partial) autocorrelation, with the penalty pushing the estimator toward a target selected from the data. This both encapsulates and differs from previous attempts at penalized estimation for autocorrelation, which shrink the estimator toward the target value of zero. Unlike the regression case, in which the least squares estimator is unbiased and shrinkage is used to reduce mean squared error by introducing bias, in the autocorrelation case the usual estimator has bias toward zero. The penalty can be chosen so that the resulting estimator of autocorrelation is asymptotically normally distributed. Simulation evidence indicates that the proposed estimators of (partial) autocorrelation tend to alleviate the bias and reduce mean squared error compared with the traditional sample ACF/PACF, especially when the time series has strong correlation.

One application of the penalized (partial) autocorrelation estimator is portmanteau tests in time series. Target and tuning parameters can be selected to improve time series Portmanteau tests–shrinking small magnitude correlations toward zero controls type I error, while increasing larger magnitude correlations improves power. Specific data based choices for target and tuning parameters are provided for general classes of time series goodness of fit tests. Asymptotic properties of the proposed test statistics are obtained. Simulations show power is improved for all of the most prevalent tests from the literature and the proposed methods are applied to data.

Another application of the penalized ACF/PACF considered in this dissertation is the optimal linear prediction of time series. We exploit ideas from high-dimensional autocorrelation matrix

estimation and use tapering and banding, as well as a regularized Durbin-Levinson algorithm to derive new predictors that are based on the penalized correlation estimators. We show that the proposed estimators reduce the error in linear prediction of times series. The performance of the proposed methods are demonstrated on simulated data and applied to data.

Dedication

I dedicate my dissertation work to my family. A special feeling of gratitude of my parents, who have always loved me unconditionally and encouraged me to pursue my dreams all the way.

I also dedicate this dissertation and give special thanks to my advisor Dr. Colin Gallagher for his unwavering support and encouragement throughout my PhD studies.

This work is also dedicated to my many friends who have encouraged me and never left my side. I will always appreciate all they have done.

Acknowledgments

Throughout the writing of this dissertation I have received a great deal of support and assistance.

First and foremost I would like to thank my PhD advisor, Dr. Colin Gallagher, whose expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I am extremely grateful for his time, academic stimulus, good advice, encouragement and enthusiasm.

I am truly grateful to my committees, Dr. Qiong Zhang, Dr. Andrew Brown and Dr. Chris McMahan, for their valuable guidance and technical support throughout my studies. They provided me with the professional advice that I needed to choose the right direction and successfully complete my dissertation.

I would also like to acknowledge and thank my colleagues from my internship at Data Support Core of Prisma Heath for their wonderful collaboration. I would particularly like to single out my supervisor at Prisma Health, Dr. Alex Ewing. I want to thank him for his support, mentor and for all of the opportunities I was given to further my research.

In addition, I would like to thank my parents for their wise counsel, sympathetic ear and tremendous understanding. They are always there for me. Finally, I am particularly thankful to my husband, Ran Huang for his encouragement and support all through my studies.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Autocorrelation and Partial Autocorrelation Function

1.1.1 Definitions and Background

The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are frequently used in time series analysis and forecasting. Autocorrelation measures the strength of the linear relationship between time series values at different points say t and s, whereas partial autocorrelation accounts for the information contained in measurements taken between time t and time s. Estimation of the (partial) autocorrelation function is fundamental in time series analysis and has applications in econometrics, finance, communications engineering, neuroscience, environmental and ecological studies, etc. They are both employed in modeling stationary time series, especially in fitting autoregressive moving average (ARMA) processes (Box and Jenkins, 1976).

Consider a stationary process $\{X_t\}$, the autocorrelation function (ACF) is defined as the function whose value at lag h is

$$\rho_X(h) = \frac{\gamma_X(h)}{\gamma_X(0)} = Corr(X_{t+h}, X_t)$$

for all $t, h \in \mathbb{Z}$. The autocovariance function (ACVF) is $\gamma(h) = E[(X_{t+h} - \mu)(X_t - \mu)]$, where $\mu = E[X_t]$. Estimating the autocorrelation function is essential for determining the degree of serial correlation in a time series. It also plays a important role in statistical inference in time series

analysis. In much of time series analysis, it is common to fit constants to estimate the mean and trend, and use the residuals to estimate autocorrelations. The most frequently used estimator is the sample autocorrelation function (sample ACF), sometimes it is called conventional autocorrelation function (Brockwell and Davis, 1991, 2002). Given the data $x_1, ..., x_n$, the estimator which we typically use for $\gamma(h)$ is the sample ACVF

$$\hat{\gamma}(h) = n^{-1} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}), \quad 0 \leq h \leq n-1$$

where $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$ is the sample mean of $\{x_t\}$. The sample ACF estimator is

$$\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \tag{1.1}$$

The autocorrelation function provides a useful measure to determine the degree of dependence among the values of a time series at different times. It is an important tool for constructing an appropriate mathematical model of the data, and optimal linear predictors for time series are determined by mean and autocorrelation parameters (Brockwell and Davis, 1991, 2002). For stationary time series with mean μ, variance $\gamma(0)$, and autocorrelation function $\rho(h)$, consider using $\{X_1, ..., X_n\}$ to predict X_{n+h}, $h > 0$. To find the best linear predictor (BLP): $P_n(X_{n+h}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i X_i$, we minimize $E[X_{n+h} - \alpha_0 - \sum_{i=1}^{n} \alpha_i X_i]^2$ and the best linear predictor is

$$P_n(X_{n+h}) = \mu + \sum_{i=1}^{n} \alpha_i (X_{n+1-i} - \mu) \tag{1.2}$$

where $\boldsymbol{\alpha}_n$ satisfies $\boldsymbol{R}_n \boldsymbol{\alpha}_n = \boldsymbol{\rho}_n(h)$ and $\alpha_0 = \mu(1 - \sum_{i=1}^{n} \alpha_i)$. Here the mean squared prediction error is $E[X_{n+h} - P_n(X_{n+h})]^2 = \gamma(0)(1 - \boldsymbol{\alpha}_n' \boldsymbol{\rho}_n(h))$ where $\boldsymbol{\alpha}_n = (\alpha_1, ..., \alpha_n)'$, $\boldsymbol{R}_n = [\rho(i-j)]_{i,j=1}^{n}$, and $\boldsymbol{\rho}_n(h) = (\rho(h), ..., \rho(h+n-1))'$. Note that if $\{Y_t\}$ is the zero mean series defined by $Y_t = X_t - \mu$, then $P_n(X_{n+h}) = \mu + P_n(Y_{n+h})$.

The partial autocorrelation function (PACF) of $\{X_t\}$ at lag h, denoted by $\alpha(h)$ is defined as the coefficient of X_t in the best linear predictor for X_{t+h} in terms of $X_{t+h-1}, ..., X_t$. For the AR(p) model $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t$, we have $\alpha(p) = \phi_p$ and $\alpha(h) = 0$, for $h > p$. Note the $\alpha(1)$ is the same as the autocorrelation at lag one, but $\alpha(h)$ for $h > 1$ can be quite different from $\rho(h)$. Another

way to define the PACF at lag h is the correlation between X_t and X_{t+h} given (conditional on) the intermediate variables $X_{t+1}, ..., X_{t+h-1}$. Therefore the PACF at lag h is

$$\alpha(h) = \phi_{hh}, \quad h \geq 1$$

where $\alpha(0) = 1$ and ϕ_{hh} is the last component of $\boldsymbol{\phi}_h = \boldsymbol{R}_h^{-1}\boldsymbol{\rho}_h$. For data $\{x_1, ...x_n\}$, the sample partial autocorrelation function (sample PACF) is given by

$$\hat{\alpha}(h) = \hat{\phi}_{hh}, \quad h \geq 1 \tag{1.3}$$

and $\hat{\alpha}(0) = 1$, where $\hat{\phi}_{hh}$ is the last component of $\hat{\boldsymbol{\phi}}_h = \hat{\boldsymbol{R}}_h^{-1}\hat{\boldsymbol{\rho}}_h$.

To determine the best linear predictor $P_n(X_{n+h})$ in terms of $\{X_1, ..., X_n\}$, the direct approach requires the determination of a solution of n linear equations, which is difficult and time consuming for large n. The Durbin-Levinson algorithm (Brockwell and Davis, 1991) processes the autocovariance $\{\gamma(1), ..., \gamma(h)\}$ or autocorrelation $\{\rho(1), ..., \rho(h)\}$ that utilize the recursive idea for $h = 1, ..., n-1$ and computes the coefficients of the optimal one-step ahead predictor at time $t > h$. The one-step predictor $P_t(X_{t+1})$ based on the h previous realizations could be used to simplify the calculation of $P_{t+1}(X_{t+2})$. For zero-mean stationary time series $\{X_t\}$, $P_t(X_{t+1}|X_t, \ldots, X_{t+1-h}) = \sum_{j=1}^{h} \phi_{hj} X_{t+1-j}$, where ϕ_{hh} is the partial autocorrelation between X_t and X_{t+h}. The coefficients $\phi_{h1}, ..., \phi_{hh}$ can be computed recursively from the equations (Brockwell and Davis, 1991)

$$\phi_{hh} = [\gamma(h) - \sum_{j=1}^{h-1} \phi_{h-1,j}\gamma(h-j)]v_{h-1}^{-1}, \tag{1.4}$$

$$(\phi_{h1}, ..., \phi_{h,h-1})' = (\phi_{h-1,1}, ..., \phi_{h-1,h-1})' - \phi_{hh}(\phi_{h-1,h-1}, ..., \phi_{h-1,1})' \tag{1.5}$$

$$v_h = v_{h-1}(1 - \phi_{hh}^2), \quad \phi_{11} = \rho(1) \text{ and } v_0 = \gamma(0). \tag{1.6}$$

This algorithm provides the solution to the Yule-Walker system of equations (1.2) in $O(h^2)$ operations, by making efficient use of the Toeplitz structure of $\boldsymbol{R}_{h+1}$.

Some recent papers focus on optimal linear prediction in a high-dimensional setting which arises when the dimension h is of the same order of magnitude as the sample size, n. In this situation, the sample autocovariance or autocorrelation matrix is not consistent (Wu and Pourahmadi, 2009). To achieve consistency, Wu and Pourahmadi (2009) and Bickel and Gel (2011) proposed banded

estimators of the sample autocovariance matrix and introduced a banding parameter l. The banded estimator keeps the first l main diagonals of the sample autocoavarince matrix intact and sets the remaining ones to zero. McMurry and Politis (2010) proposed a tapered and banded estimator in which the $2l+1$ subdiagonals of the sample autocovariance or autocorrelation matrix are maintained the same and the remaining entries are shrunk to zero gradually. McMurry and Politis (2015) introduced a consistent estimator of the $n \times n$ autocovariance matrix to construct an estimator of the optimal, full-length coefficient vector. They developed the methodology for optimal linear prediction of a stationary time series based on the resulting autocovariance estimates. However, the tapered and banded matrices are not positive definite. Proietti and Giovannelli (2018) proposed a Durbin-Levinson regularized estimator that receives as input the banded and tapered sample PACF. This estimator is consistent and positive-definite.

Time series goodness of fit tests are primarily based on estimates of the ACF and PACF of the residuals from a fitted model. Many authors have proposed goodness of fit tests to check the adequacy of the fitted ARMA model. The commonly used test statistics are those proposed by Box and Pierce (1970), Ljung and Box (1978), Li and Mak (1994), Monti (1994), Peña and Rodríguez (2002, 2006), Mahdi and McLeod (2012) and Fisher and Gallagher (2012) using the sample ACF or PACF. McLeod and Li (1983a) introduced a portmanteau test using the squared-residuals autocorrelations for detecting nonlinearity (ARCH). To test for neglected nonlinearity in time series, several improved portmanteau statistics have been proposed (Li and Mak, 1994; Rodríguez and Ruiz, 2005; Peña and Rodríguez, 2002, 2006). Fisher and Gallagher (2012) proposed a portmanteau statistics that are weighted sum of the squares of the sample autocorrelation coefficients. The weighted tests are modified to check for nonlinearity and the adequacy of a fitted nonlinear model. Simulation studies show that the weighted goodness of fit tests tend to have higher power than traditional tests.

Also, when conducting inference with autocorrelated data, determining appropriate corrections to standard errors is a common problem in time series econometrics (Berk, 1974; Newey and West, 1987). In many cases the goal is to make inference in terms of regression parameters without assuming a specific model for the correlation structure. The heteroscedasticity and autocorrelation consistent (HAC) methods (Andrews, 1991; Müller, 2014) estimate the standard error of the estimated mean (regression) parameters using a weighted linear combination of estimated autocorrelations. HAC estimators considered perform poorly in an absolute sense with high amount

of correlation. The procedures mentioned above are dependent on accurate estimates of the degree of correlation. However, it has long been known that the sample ACF and PACF are biased and tend to underestimate the magnitude of the correlation.

1.1.2 Estimators of Autocorrelation Function

In fact, several alternative estimators had been proposed and widely studied in the literature (Bartlett, 1946; Kendall, 1954; Marriott and Pope, 1954; White, 1961; Andrews, 1993). Much of the motivation behind these studies was to reduce the bias of the conventional estimator. The main objective of this section is to describe six autocorrelation estimators, presenting the expressions for their calculus. Note that the conventional (sample), ordinary least, modified , and Fuller estimators have the general form for multi lags, while the others are estimators of lag-one autocorrelation function.

Conventional estimator: the sample ACF, is widely used in most of the texts and in social and behavioural science studies. It is also known as the Yule-Walker method for ARMA models. The estimator is defined by the expression in (1.1). Kendall and Ord (1990) considered the bias of conventional autocorrelation estimation under an AR(1) model. The bias is $-\frac{(1+4\rho)}{n}$ for long series, where ρ is the autoregressive parameter and n is the series length.

Least squares estimator: another widely used estimator is the least square estimator of the autocorrelation. The ordinary least squares (LS) estimator can be expressed in the following manner:

$$\hat{\rho}_{LS}(h) = \frac{\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n-h}(X_t - \bar{X})^2}, \quad h = 1, 2, ..., n-1 \tag{1.7}$$

When we consider the lag-one autocorrelation estimator, this is an estimator of the autoregressive parameter in the first order autoregressive model, because $\phi_1 = \rho(1)$. In general, if the model is written as $X_t = \phi_h X_{t-h} + \epsilon_t$, the least squares solution for the estimator of ϕ_h can be written as $(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\boldsymbol{y}$, where $\boldsymbol{x} = \{X_i - \bar{X}\}_{i=1}^{n-h}$, $\boldsymbol{y} = \{X_j - \bar{X}\}_{j=h+1}^{n}$, and $\bar{X} = \sum_{t=1}^{n} X_t/n$. This estimator has the same number of terms in the numerator and the denominator $n-h$. Huitema and McKean (1991) also investigated the properties of least square estimator in an extensive Monte Carlo study. They found that the least square estimator performs similarly to the conventional estimator, by comparing the bias and mean squared errors. However, It is still highly biased, especially when

$|\rho(h)|$ tends to one.

Modified estimator: the modified estimator (Orcutt and Irwin, 1948) consists of a linear modification of the conventional estimator. This estimator is proposed in the following formula,

$$r^*(h) = \frac{n}{n-h} \frac{\sum_{t=1}^{n-h}(X_t - \bar{X})(X_{t+h} - \bar{X})}{\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{1.8}$$

It is defined as the sample ACF estimator multiplied by a term greater than one $\frac{n}{n-h}\hat{\rho}(h)$, where $\hat{\rho}(h)$ is the sample ACF estimator at lag h and n is the sample size. The bias of the modified estimator approximates $-\frac{(1+3\rho)}{n}$ for long series, which is less than that of the conventional estimator. For the conventional estimator, the number of terms in numerator is fewer than terms in denominator. This discrepancy leads to conventional estimates that are closer to zero, which is called "the squeezing effect" (Huitema and McKean, 1991). So multiplying by some term greater than one will help to remove the effect of the discrepancy in the number of terms.

Fuller's estimator: Fuller (1976) attempted to correct the sample estimator's bias, especially for short series. The Fuller estimator has the following expression:

$$r^f(h) = \hat{\rho}(h) + \frac{n-h}{(n-1)^2}(1 - \hat{\rho}(h)^2). \tag{1.9}$$

C statistic: Young (1941) proposed the C statistic to determine whether or not data series are random. The C statistic is an estimator of lag-one autocorrelation (DeCarlo and Tryon, 1993). It add a factor to $\hat{\rho}$ to compensate the bias of the $\rho(1)$ estimator as follows,

$$C = 1 - \frac{\sum_{t=1}^{n-1}(X_t - X_{t+1})^2}{2\sum_{t=1}^{n}(X_t - \bar{X})^2} = \hat{\rho}(1) + \frac{(X_n - \bar{X})^2 + (X_1 - \bar{X})^2}{2\sum_{t=1}^{n}(X_t - \bar{X})^2} \tag{1.10}$$

Translated estimator: Huitema and McKean (1991) proposed the r_1^+ estimator as follows,

$$r_1^+ = \hat{\rho}(1) + \frac{1}{n}. \tag{1.11}$$

Since it performs a translation over the sample estimator by adding a bias correction part n^{-1}, it is referred as the translated estimator. The bias of r_1^+ is approximately $-(4\rho)/n$.

Additionally, there are other estimators, such as, exact estimator Kendall (1954), the σ-recursive estimator, the maximum likelihood estimator (Kendall and Ord, 1990). Anderson (1942) investigated a cyclic estimator for different lag autocorrelation. This estimator has one extra term in the numerator $(X_n - \bar{X})(X_{n+1} - \bar{X})$, where $X_{n+1} = X_1$. Huitema and McKean (1994) investigated two new autocorrelation estimators, r_{F1} and r_{F2}, which are far less biased than the conventional estimator, but their error variances increase. Arnau and Bono (2001) proposed a first-order ACF estimator by analyzing the function of the empirical bias with the polynomial regression for different sample sizes and correcting by the absolute value of the polynomial fitting model. Although bias in autocorrelation estimation has been recognized and various alternative estimators have been proposed in several decades, there is no estimator of autocorrelation that can solve the underestimation issue of sample ACF for all positive and negative autocorrelations in stationary ARMA processes with different lags. The limitations of autocorrelation estimators in the literature are: (1) most of the studies only focus on the lag-one autocorrelation in time series; (2) the bias corrected estimators perform better than sample ACF for positive correlations, but worse for negative ones; (3) these estimators have higher mean squared error (MSE); (4) they are not positive definite estimators. Our work will use penalized regression methods to estimate (partial) autocorrelation to reduce the bias and improve the performance in terms of mean squared error for the stationary time series.

1.2 Penalized M-estimators

Penalized objective function estimation for regression models has been extensively studied (Hoerl and Kennard, 1970; Tibshirani, 1996; Fan and Li, 2001; Zou, 2006). In the regression context least squares provides unbiased estimators of the regression parameters. Penalizing the objective function based on the distance of regression parameters from the origin introduces bias while reducing variance, which can reduce MSE by "shrinking" the parameter estimates toward zero. A general form of the penalized least square estimator is to minimize

$$O(data; \boldsymbol{\beta}) + ||\Lambda\boldsymbol{\beta}||,$$

where O is an objective function, $\boldsymbol{\beta} = (\beta_1, ...\beta_p)$ are the regression parameters, Λ is a diagonal matrix of non-negative of tuning parameters, and $||\cdot||$ denotes a distance between the regression

parameters β and the origin. Typically, the first part of the objective function is the sum of least squares errors $||Y - X\beta||^2$, where X, Y are from data. The penalty function takes different forms for different methods.

In the background of regularization, Hoerl and Kennard (1970) introduced the ridge regression which can provide parameter estimates with small variance for data and obtain better performance in prediction. Ridge regression minimize an objective function containing a least squares error term and a L_2 penalty term which is defined as

$$||\Lambda\boldsymbol{\beta}||_{Ridge} = \lambda \sum_{i=1}^{p} \beta_i^2$$

where $\lambda \geq 0$ is the tuning parameter that controls the amount of shrinkage. It shrinks the regression coefficients by imposing a penalty on their sizes. Our work will consider the L_2 norm in the penalty term.

The Lasso method is a L_1 penalized least squares method which is proposed by Tibshirani (1996). The penalty term of this method is referred to as a L_1 penalty:

$$||\Lambda\boldsymbol{\beta}||_{Lasso} = \lambda \sum_{i=1}^{p} |\beta_i|$$

where λ is a non-negative regularization parameter whose role is to balance prediction accuracy and sparsity. When the L_1 penalty is applied, many regression coefficients are shrunk to zero and a few other regression coefficients are shrunk comparatively little. The Lasso method can be used in the situation when there are more predictors p than subjects n in a study. It also corresponds to a convex minimization procedure. The Elastic Net method (Zou, 2006) is proposed to improve the Lasso when the predictors are highly correlated. To compensate for the correlation between predictors, the penalty of the Elastic Net is

$$||\Lambda\boldsymbol{\beta}||_{ElasticNet} = \lambda_1 \sum_{i=1}^{p} |\beta_i| + \lambda_2 \sum_{i=1}^{p} \beta_i^2$$

for given tuning parameters λ_1 and λ_2. The minimization problem is also convex. The penalty tends to result in all small but nonzero regression coefficients. The Elastic Net estimator shares the sparsity properties of the Lasso estimator and leads to more accurate prediction of the response than Lasso estimator when predictors have high correlation. To compare among the estimators of OLS,

ridge regression, Lasso, Elastic Net, it is well known that the ridge estimator is a constant scaled OLS estimator, the Lasso estimator is a constant translated OLS estimator and elastic net includes ridge type shrinkage and lasso type thresholding.

Fan and Li (2001) proposed the SCAD method in a general parametric framework for efficient estimation and variable selection. This method applies a specially designed penalty function, the smoothly clipped absolute deviation (SCAD) (Fan and Li, 2001; Fan, 2004). The continuous penalty function is defined by

$$||\Lambda\boldsymbol{\beta}||_{SCAD} = \lambda I(\beta \leq \lambda) + \frac{(a\lambda - \beta)_+}{(a-1)\lambda} I(\beta > \lambda)$$

for some $a > 2$ and $\theta > 0$, where $a \approx 3.7$ as suggested by Bayesian risk analysis for SCAD. It has been showed that the lasso doesn't satisfy the oracle properties of a good procedure (Fan and Li, 2001). Zou (2006) proposed the adaptive lasso for simultaneous parameter estimation and variable selection. The adaptive lasso enjoys the oracle properties using the adaptively weighed L_1 penalty which is

$$||\Lambda\boldsymbol{\beta}||_{Adaptive} = \lambda \sum_{i=1}^{p} \hat{w}_i |\beta_i|$$

where $\boldsymbol{w}$ is a known weight vector with elements which are data-dependent. For example, we can use $\hat{\beta}_{ols}$, and the weights $\hat{w}_i = 1/|\hat{\beta}_i^{\gamma}|$ with $\gamma > 0$. The adaptive lasso shrinkage also leads to a near-minimax-optimal estimator and enjoys the computational advantage of the Lasso.

In the case of estimating autocorrelation, the finite sample bias is toward zero. To alleviate this bias and reduce MSE, we propose penalizing correlation values based on the distance between the correlation and a target $\boldsymbol{\rho}_\tau$. Consider a general penalized objective function:

$$O(data; \boldsymbol{\rho}) + ||\Lambda(\boldsymbol{\rho} - \boldsymbol{\rho}_\tau)||, \tag{1.12}$$

where the penalty term moves the resulting penalized estimator toward the target. The target values and tuning parameters can depend on both the data and the applications.

1.3 Cross validation Methods with Dependent Data

Penalized least squares minimization requires us to choose tuning parameters. Cross validation (CV) (Stone, 1974; Arlot and Celisse, 2010) is a popular method for parameter estimation. The fundamental principle of cross validation is to partition data once or multiple times in order to estimate the risk of each algorithm: a portion of the data (the training sample) is used to train each algorithm, while the remainder (the validation sample) is used to estimate the method's risk. The tuning parameter with the smallest estimated risk is then chosen by cross validation. When data are identical and independent distributed (IID), the training sample is independent from the validation sample. Then cross validation avoids overfitting compared to other methods. The popularity of the cross validation method mostly comes from the common strategy which is data splitting.

When data are dependent, the cross validation methods break down since the validation and training samples are no longer independent. It is often unclear which is the best way to evaluate models in time series cases. Some studies demonstrate cases where traditional cross validation fails in time series contexts (Hart and Wehrly, 1986; Chu and Marron, 1991). Opsomer et al. (2001) showed that if the autocorrelation of the error is high, and the approach overfits the data, traditional cross validation underestimates bandwidths in a kernel estimator regression framework. The bandwidth selectors developed for independent observations will not produce good bandwidths if the observations are dependent. For example, cross validation will provide small bandwidths if the data are positively correlated, resulting in rough kernel estimations of the regression function. On the other hand, cross validation will create large bandwidths if the data are negatively correlated, resulting in oversmooth kernel estimations of the regression function. Therefore cross validation must be modified. Several cross validation techniques for dependent data have been investigated extensively in the literature (Györfi et al., 1989; Burman and Nolan, 1992; Burman et al., 1994; Hyndman, 2014; Bergmeir et al., 2018). In what follows we briefly introduce some approaches.

h-block CV method: Let $\boldsymbol{Y} = \{Y_1, ..., Y_n\}$ be a time series. The traditional K-fold CV randomly removes K numbers out of the vector $\boldsymbol{Y}$. But the errors in the training and test tests are correlated in the dependent setting. The h-block CV proposed by Burman et al. (1994) which reduces the training set by removing the h observations preceding and following the observations in the test set. h is a fixed fraction of the sample size. It excludes any observations that could be considered dependent and only includes those that can be regarded independent. The limitation of

this method is inefficient use of all of the available data. Since the h-block CV is not asymptotically optimal. Racine (2000) proposed a modification of h-block method, named hv-block cross validation, which is consistent for general stationary observations.

K-fold CV method for AR models: Although cross-validation is sometimes not valid for time series cases, it is possible to use the traditional K-fold CV if the models considered have uncorrelated errors for purely autoregressive models (Bergmeir et al., 2018). Consider an AR model with a fixed lag order h, the embedded time series with order h can be written as a matrix as follows,

$$\begin{bmatrix} Y_1 & Y_2 & \dots & Y_h & Y_{h+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{t-h} & Y_{t-h+1} & \dots & Y_{h-1} & Y_t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Y_{n-h} & Y_{n-h+1} & \dots & Y_{h-1} & Y_n \end{bmatrix}.$$

The matrix is used as the input of a regression where each row is the form of $[\boldsymbol{X}_t^{'}, Y_t]$ and $\boldsymbol{X}_t^{'} = (Y_t, Y_{t-h+1}, ..., Y_{h-1})^{'}$. Then the standard K-fold CV method can be applied. The data is randomly partitioned into K groups. Let $P = \{P_1, ..., P_k\}$, P_k is taken as test set and the remaining groups $P_{k^-} = \cup_{j \neq k} P_j$ are taken as training set. Usually we use $k = 5, 10, n - h$. The idea of this method is to leave out the entire set of rows at t in the matrix rather than removing the h observations to decrease the training set. They showed empirically that K-fold CV performs better than out of sample evaluation and non-dependent CV method.

Leave-one-out CV (LOOCV) method: This method is a special case of the k-fold CV method. With the fixed lag order of the model h, the number of fold is $k = n - h$ which is the number of rows in the embedded matrix. Each fold includes only one row of the matrix. While LOOCV can be very time consuming to implement in general, it is fast to compute LOOCV for linear models.

Time Series CV (tsCV) method: For time series forecasting, Hyndman (2014) introduced time series cross-validation which is based on one-step forecasts. It fits a model to the data $Y_1, ..., Y_n$ and let Y_{t+1} be the test data for $t = k + 1, ..., n - 1$ where k is the minimum number of observations needed for fitting the model. However, there are limitations using cross-validation for time series, such as, inefficient use of the available data.

In Chapter 2, we will use LOOCV to determine the optimal value for a tuning parameter in the penalized estimation of autocorrelation, as its value regulates how much we "shrink" to our target correlations.

1.4 Dissertation Organization

The remainder of this dissertation is organized as follows.

In Chapter 2, we propose the estimation of autocorrelation and partial autocorrelation using penalized regression methods for stationary time series. We prove the asymptotic properties of the penalized autocorrelation estimator, and introduce a way to ensure positive definiteness in finite samples. Some approaches to select tuning parameters will be discussed. With respect to bias and mean square error, we conduct Monte Carlo study on the performance of the penalized ACF/PACF estimator. The proposed methodology shows promising results for stationary ARMA(p, q) processes.

In Chapter 3, we review portmanteau statistics for stationary time series goodness of fit testing, and then propose a novel penalized M-estimator of (partial) autocorrelation, with the target and tuning parameters improving time series Portmanteau tests. We develop the penalized statistics for fitted ARMA processes and detecting nonlinearity, and prove asymptotic properties. Simulations show the type I error is controlled and power is improved for traditional goodness of fit tests. The proposed methods are applied to real world time series data.

In Chapter 4, we proposed a penalized taper and banded estimator of (partial) autocorrelation matrices and a Durbin-Levinson regularized estimator of penalized autocorrelation matrices for the stationary random process in prediction. A computational algorithm based on the proposed estimators will be introduced. Simulation study and data application on the performance of the proposed estimator for prediction will be conducted.

In Chapter 5, some general conclusions and discussion are provided.

Appendix A reviews the autocorrelation functions of AR(p), MA(q) and ARMA(p,q) models. Appendix B shows the function documentation for the R package "PenalizedPortTest" from Chapter 3.

Chapter 2

Penalized (Partial) Autocorrelation Function Estimation

2.1 Introduction

The estimation of the (partial) autocorrelation function plays a central role in time series analysis. The autocorrelation and partial autocorrelation function (Brockwell and Davis, 1991, 2002) can be used for model identification, for fitting AR models using the Yule-Walker equations, for time series goodness of fit (Box and Pierce, 1970; Ljung and Box, 1978; Fisher and Gallagher, 2012), and for predicting future time series values (Brockwell and Davis, 1991). The bias issue of the conventional estimators, the sample ACF in (1.1) and PACF in (1.3), is well known. Back in 1948, Moran (1948) studied the theoretical bias in all serial correlations of a random series, which is not necessarily normal. The bias using derivations from the sample mean is $-(n-1)^{-1}$ for both circular and non circular definitions. Marriot and Pope (1954) pointed out two sources of the bias. Firstly, the autocorrelation will be biased if the true mean of the series is known, except when the corresponding autocorrelations are zero, since the estimate variance and correlation are not independently distributed. The second source is that the mean is usually estimated from the sample in practice and this introduces more bias.

In practice, several investigations (Huitema and McKean, 1991; DeCarlo and Tryon, 1993; Arnau and Bono, 2001; Solanas et al., 2010) have carried out Monte Carlo simulation comparisons of

autocorrelation estimators. Huitema and McKean (1991) pointed out that the conventional estimator has larger empirical and theoretical bias at $\rho = 0.9$ and $\rho = -0.9$. The results are consistent with Kendall's (1954) conjecture that autocorrelation are probably satisfactory for values of ρ near zero, but they are of doubtful validity for ρ near to unity.

Figure 2.1 illustrates the theoretical and empirical bias in sample ACF of autocorrelative coefficient for different model cases with one lag $h = 1$ or 2. Samples were generated from the ARMA models with different sample sizes ($n = 20, 50, 100$) and true autocorrelation values $[-0.99(0.05)0.99]$ or $[-0.5, 0.5]$ for MA(1) case. It is of interest to notice that the estimates are significantly highly biased near boundaries, although the bias decreases as the sample size increases. The bias increases with the magnitude of the correlation. Note that the bias values in AR(1) obtained with sample autocorrelation are the same as the results of Huitema and McKean (1991) who used the International Mathematical and Statistical Libraries (1989), and those of Arnau and Bono (2001) who used the MATLAB program.

In this chapter, we consider estimating (partial) autocorrelation of stationary time series using penalized techniques. Under some conditions, we show that the new estimator has the same asymptotic distribution as the sample autocorrelation function. While bias correction formulas can be calculated for specific assumed models, no general bias correction formulas are available. Unlike the regression case, in which the least squares estimator is unbiased and shrinkage is used to reduce mean squared error by introducing bias, in the autocorrelation case the usual estimator has bias toward zero. We thus choose penalties that penalize for underestimating the magnitude of the correlation and thereby reduce both bias and MSE.

The rest of the chapter is structured as follows. Section 2.2 introduces a new (partial) autocorrelation estimator using penalized regression methods, and presents its asymptotic properties of the proposed estimator and positive definiteness. The selection of the tuning parameters is discussed in Section 2.3. Performance of proposed ACF/PACF estimators is assessed via simulation for linear processes in Section 2.4. The results are more satisfactory for high and low correlations and comparative with the sample ACF/PACF estimator for moderate correlations.

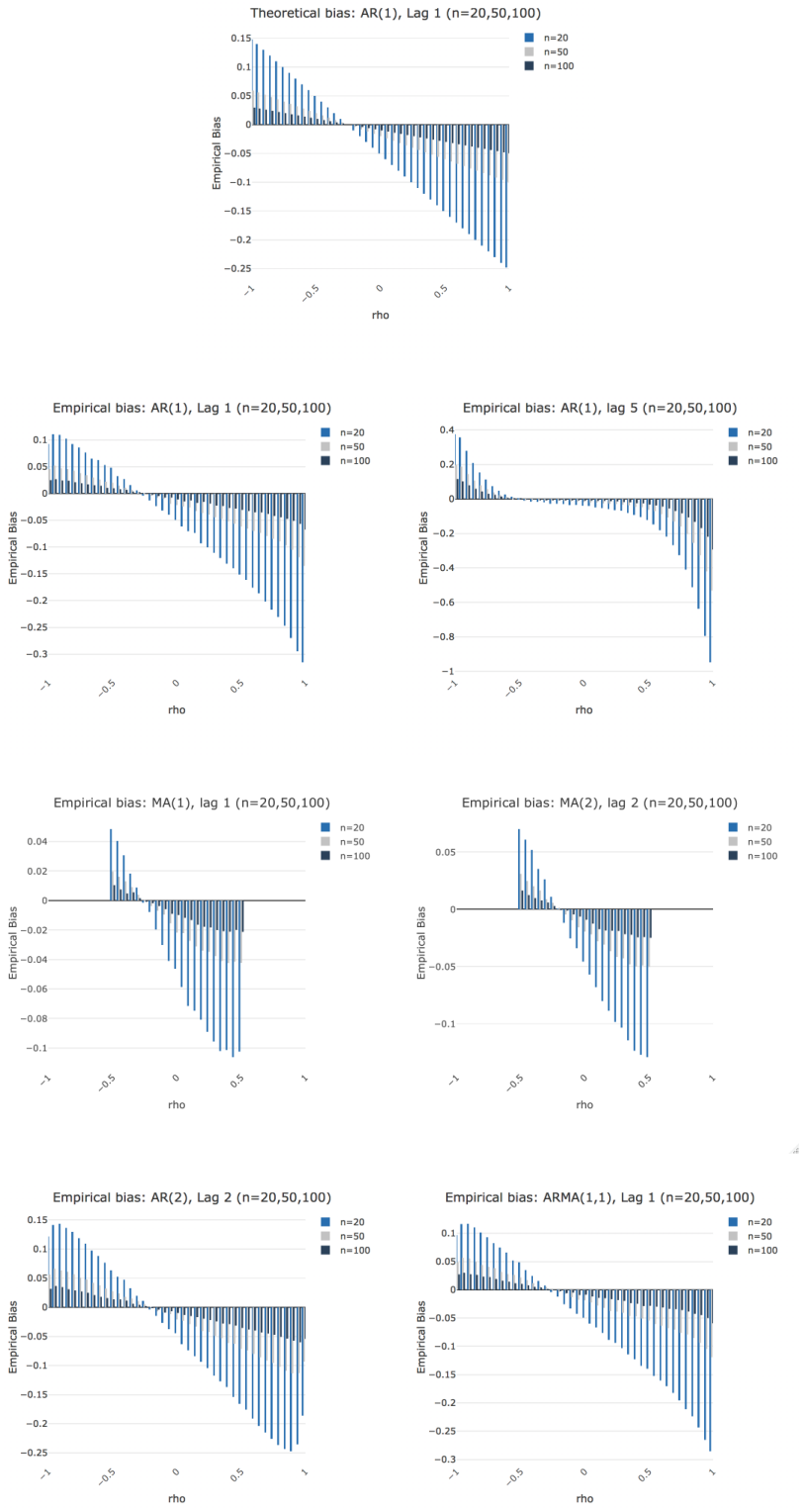


Figure 2.1: Empirical and Theoretical Bias of Sample ACF

2.2 Penalized Estimator and Properties

Let $\{X_t, t \in \mathbb{Z}\}$ be a stochastic process with mean μ which is described by

$$X_t - \mu = \sum_{j=0}^{\infty} \psi_j Z_{t-j}, \quad t \in \mathbb{Z} \tag{2.1}$$

where $\{Z_t\}$ is iid with $E[Z_t] = 0$, $E|Z_t|^2 < \infty$, and some sequence $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$. Such linear processes are also called moving-average (MA) processes. It is well known (Brockwell and Davis, 1991) that the causal autoregressive (AR) or ARMA processes are representable as the linear processes (MA processes) $\{X_t\}$ defined in (2.1). In this work, our main objective is to propose a new estimation of the autocorrelation function ρ (or partial autocorrelation function) under linear process models and we consider only the solutions such as $|\rho(h)| \leq 1$ for $h \in \mathbb{Z}^+$. For a time series with expectation $E[X_t] = \mu$ in (2.1) and autocorrelation function

$$\rho(h) = E(X_t - \mu)(X_{t-h} - \mu)/E(X_t - \mu)^2 \quad \text{for} \quad h = 0, \pm 1, \pm 2, \ldots$$

It is easy to see that

$$\rho(|i-j|) = \arg\min_{\rho} E[(X_i - \mu) - \rho(X_j - \mu)]^2, \quad i, j \in \mathbb{Z} \tag{2.2}$$

Then based on this idea, one can set up a regression model as follows,

$$X_i - \mu = \rho(|i-j|)(X_j - \mu) + \epsilon_{ij}, \quad i, j \in \mathbb{Z} \tag{2.3}$$

where ϵ_{ij} are independent standard normal variables. Given the observations $\{x_1, x_2, ..., x_n\}$ from the stationary time series $\{X_t\}$ discussed above, we can estimate μ with $\bar{x} = n^{-1} \sum_{t=1}^{n} x_t$. A variety of correlation estimators have been proposed in the literature and most of these can be associated with an objective function which approximates (2.2), with the primary differences between the estimators being how the boundary values (t near 1 and n) are handled. Many time series analysts prefer the sample ACF, since the resulting function is guaranteed to be non-negative definite. The sample ACF at lag h is calculated as (1.1). Also, the autocorrelation $\rho(|i-j|)$ for $1 < i, j < n$ can be obtained by the least square methods. Chang and Politis (2016) constructed an autoregressive

scatterplot associated with the pairs $\{(x_t - \bar{x}, x_{t+h} - \bar{x})\}$ where $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$ and run the linear regression on the pairs to get a robust estimate of autocorrelation. Here, we can use different pairs to generate typical estimators of the autocorrelation function.

Assume we have the augmented data $x_1^*, x_2^*, ..., x_n^*, x_{n+1}^*, ..., x_{2n-1}^*$ where $x_t^* = x_t$ for $t = 1, 2, ..., n$ and $x_t^* = \bar{x} = n^{-1}\sum_{t=1}^{n} x_t$ for $t = n+1, ..., 2n-1$. The least square fit of the autocorrelation at lag h, $h < n$ can be obtained as

$$\hat{\rho}_{LS}(h) = \arg\min_{\rho(h)} \sum_{t=1}^{n-h} [(x_{t+h}^* - \bar{x}) - \rho(h)(x_t^* - \bar{x})]^2$$

where $\bar{x} = \bar{x^*}$. Then, we have the least square (LS) estimator $\hat{\rho}_{LS}(h)$

$$\hat{\rho}_{LS}(h) = \frac{\sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n-h} (x_t - \bar{x})^2} \tag{2.4}$$

For the sample ACF, the estimator can be expressed as

$$\hat{\rho}(h) = \arg\min_{\rho(h)} \sum_{t=1}^{n} [(x_{t+h}^* - \bar{x}) - \rho(h)(x_t^* - \bar{x})]^2$$

Then,

$$\hat{\rho}(h) = \frac{\sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x})}{\sum_{t=1}^{n} (x_t - \bar{x})^2} \tag{2.5}$$

Here, the least square estimator is almost identical to the sample ACF for $\frac{h}{n}$ small (Chang and Politis, 2016). In the stationary case, it makes sense to minimize the sum of squares of the forward and backward errors (Ulrych and Clayton, 1976; Tuan, 1992) and yield a forward-backward least square (FBLS) estimator $\hat{\rho}_{FBLS}(h)$

$$\hat{\rho}_{FBLS}(h) = \arg\min_{\rho(h)} \sum_{t=1}^{n-h} \left\{ [(x_{t+h}^* - \bar{x}) - \rho(h)(x_t^* - \bar{x})]^2 + [(x_t^* - \bar{x}) - \rho(h)(x_{t+h}^* - \bar{x})]^2 \right\} \tag{2.6}$$

Then we have

$$\hat{\rho}_{FBLS}(h) = \frac{\sum_{t=1}^{n-h}(x_{t+h}-\bar{x})(x_t-\bar{x})}{\frac{1}{2}\sum_{t=1}^{h}(x_t-\bar{x})^2 + \sum_{t=h+1}^{n-h}((x_t-\bar{x})^2 + \frac{1}{2}\sum_{t=n-h+1}^{n}(x_t-\bar{x})^2} \tag{2.7}$$

Noting that estimating $\rho(h)$ for $h < n$ by conventional method, least squares method or forward backward least square method leads to a biased solution. We apply a penalized least squared technique with constraints to the autocorrelation estimation problem.

In this work, we extend the model in (2.3) to a more general form that can represent the sample ACF, the least square ACF and the forward-backward ACF. We propose a new estimator $\tilde{\boldsymbol{\rho}}$ of the autocorrelation $\boldsymbol{\rho} = (\rho(1), ..., \rho(n-1))^{'}$ as, let $Y_t = x_t - \bar{x}$ which is a time series with sample mean of zero, and the corresponding augmented data set is Y_t^* where $Y_t^* = Y_t$ for $t = 1, 2, ..., n$ and $Y_t^* = 0$ for $t = n+1, ..., 2n-1$. A general version of the penalized model can be expressed as

$$\begin{aligned}\tilde{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \sum_{h=1}^{n-1} \Big\{ c_f \sum_{t=1}^{n-h} \left[Y_{t+h}^* - \rho(h) Y_t^*\right]^2 + c_s \sum_{t=n-h+1}^{n} \left[Y_{t+h}^* - \rho(h) Y_t^*\right]^2 \\ + c_b \sum_{t=1}^{n-h} \left[Y_t^* - \rho(h) Y_{t+h}^*\right]^2 \Big\} + \sum_{h=1}^{n-1} v_h \mathbf{p}_\lambda(\rho(h))\end{aligned} \tag{2.8}$$

on $\{\rho : |\rho(h)| \leq 1, h = 1, ...n-1\}$, where $0 \leq c_f, c_s, c_b \leq 1$. We introduce the generalized penalty $||\Lambda(\boldsymbol{\rho} - \boldsymbol{\rho}_\tau)|| = \sum_{h=1}^{n-1} v_h \mathbf{p}_\lambda(\rho(h))$ where v_h is the function of $\{Y_t\}$ and $\mathbf{p}_\lambda(\rho(h))$ are the penalty functions which depend on λ. The specific choices of c_f, c_s, c_b and v_h are given in Table 2.1. Other unpenalized ACF estimators might be expressed similarly. In this work, we consider the above three ACF estimators.

Table 2.1: Weights for the three types of ACF estimators in model (2.8)

Unpenalized Estimators $\hat{\boldsymbol{\rho}}^*$	c_f	c_s	c_b	v_h
$\hat{\boldsymbol{\rho}}$	1	1	0	$\sum_{t=1}^{n} Y_t^2$
$\hat{\boldsymbol{\rho}}_{LS}$	1	0	0	$\sum_{t=1}^{n-h} Y_t^2$
$\hat{\boldsymbol{\rho}}_{FBLS}$	1	0	1	$\frac{1}{2}\sum_{t=1}^{h} Y_t^2 + \sum_{t=h+1}^{n-h} Y_t^2 + \frac{1}{2}\sum_{t=n-h+1}^{n} Y_t^2$

The unpenalized estimate is obtained via minimizing the whole first term which is equivalent to $||\hat{\rho}^*(h) - \rho(h)||$ where $\hat{\rho}^*(h) \in \{\hat{\rho}(h), \hat{\rho}_{LS}(h), \hat{\rho}_{FBLS}(h)\}$. Thus the form of the penalized least

square is

$$O_h(\rho(h)|Y_1^*,\ldots,Y_n^*) = \sum_{i,j}(Y_j^* - Y_i^*)^2 + \sum_{h=1}^{n-1}(\hat{\rho}^*(h) - \rho(h))^2 v_h + \sum_{t=1}^{h} v_h \mathbf{p}_\lambda(\rho(h))$$

where $1 \le i < j \le n$ for $\hat{\rho}^* = \hat{\rho}_{LS}$, $1 \le i < j \le n+h$ for $\hat{\rho}^* = \hat{\rho}$ and $1 \le i, j \le n$ for $\hat{\rho}^* = \hat{\rho}_{FBLS}$. The minimization problem of the above equation is equivalent to minimizing componentwise. This leads us to consider the following least square problem for each lag $h = 1, ..., n-1$

$$(\hat{\rho}^*(h) - \rho(h))v_h + v_h \mathbf{p}_\lambda(\rho(h))$$

The penalty functions $\mathbf{p}_\lambda(.)$ are not necessarily the same for all h. For example, we may wish to shrink ρ to zero for weak correlation and increase ρ for moderate and strong correlation. We assume that there is a target autocorrelation ρ_τ to achieve our goal. So $\mathbf{p}_\lambda(.)$ may be allowed to depend on ρ_τ and is defined by

$$||\Lambda(\boldsymbol{\rho} - \boldsymbol{\rho}_\tau)|| = \mathbf{p}_\lambda(\rho(h)) = \sum_{s=1}^{S} \lambda_{n,h}^s p_s(\rho(h)) = \sum_{s=1}^{S} \lambda_{n,h}^s (|\rho_{\tau,s}| - sgn(\hat{\rho}^*(h))\rho(h))^2 \tag{2.9}$$

Here we assume the proposed estimator has the same sign as the unpenalized ACF estimator $\hat{\rho}^*$ and $\lambda_{n,h} \ge 0$. For example, we consider $S = 3$ and the target function is given by

$$|\rho_\tau(h)| = \begin{cases} 0 & |\hat{\rho}^*(h)| \le \ell_h \\ f(|\hat{\rho}^*(h)|) & \ell_n < |\hat{\rho}^*(h)| \le u_h \\ 1 & |\hat{\rho}^*(h)| > u_h \end{cases} \tag{2.10}$$

for some thresholds $u_h \ge \ell_n > 0$. The proposed method works best for a strongly correlated time series when $\rho_\tau = \pm 1$. Moreover, $|\rho|$ shrinks to zero when $\rho_\tau = 0$. Since $\hat{\rho}^*(h)$ will be in one of the intervals, we can assume $\lambda_{n,h}^s = \lambda_{n,h}$ for $s = 1, .., S$ which depends on n and h.

The general form of **penalized autocorrelation estimator** can be written as,

$$\begin{aligned}\tilde{\rho}(h) &= \frac{\sum_{t=1}^{n-h} Y_t Y_{t+h} + sgn(\hat{\rho}^*(h))|\rho_\tau(h)|\lambda_h v_h}{v_h + \lambda_h v_h} \\ &= \frac{\hat{\rho}^*(h) + sgn(\hat{\rho}^*(h))|\rho_\tau(h)|\lambda_h}{1+\lambda_h} \\ &= w_h sgn(\hat{\rho}^*(h))|\rho_\tau(h)| + (1-w_h)\hat{\rho}^*(h) \end{aligned} \tag{2.11}$$

where $w_h = \frac{\lambda_h}{1+\lambda_h}$, $\hat{\rho}^* \in \{\hat{\rho}, \hat{\rho}_{LS}, \hat{\rho}_{FBLS}\}$ and the corresponding v_h as in table (2.1). $\hat{\boldsymbol{\rho}}$, $\hat{\boldsymbol{\rho}}_{LS}$ and $\hat{\boldsymbol{\rho}}_{FBLS}$ are special cases of the penalized estimator when $w_h = 0$. The estimator is a convex combination of the unpenalized estimator and the target:

$$\tilde{\rho}(h) = w_h \rho_\tau(h) + (1-w_h)\hat{\rho}(h), \tag{2.12}$$

Remark 1. Assume the size n is fixed, then we have

1. $|\tilde{\rho}(h)|$ decreases in λ_h when $\hat{\rho}(h) \leq \ell_h$; $|\tilde{\rho}(h)|$ increases in λ_h when $\hat{\rho}(h) > \ell_n$ and $\rho_\tau \neq \hat{\rho}^*$.

2. $\tilde{\rho}(h)$ is bounded, i.e., $\lim_{\lambda \to \infty} |\tilde{\rho}(h)| = 1$.

The following theorem states the asymptotic normality of the proposed estimator. We use $\Rightarrow$ to denote convergence in distribution.

Theorem 1 Let $\hat{\boldsymbol{\rho}}' = (\hat{\rho}(1), \ldots, \hat{\rho}(h))$, $\tilde{\boldsymbol{\rho}}' = (\tilde{\rho}(1), \ldots, \tilde{\rho}(h))$, $\boldsymbol{\rho}_\tau' = (\rho_\tau(1), \ldots, \rho_\tau(h))$ be a vector of target values, $\boldsymbol{\lambda}' = (\lambda_1, \ldots, \lambda_h)$ be a vector of tuning parameters, and $a_n(\hat{\boldsymbol{\rho}} - \boldsymbol{\rho})$ converge in distribution to L. If $a_n(\hat{\rho}(h) - \rho_\tau(h)) \to 0$ in probability, then $a_n(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho})$ converges to L in distribution.

Remark 2. If $\{X_t\}$ is the stationary process $X_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j Z_{t-j}$, $\{Z_t\} \sim IID(0, \sigma^2)$, where $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ and $E[Z_t]^4 < \infty$, and $\sqrt{n}\lambda_h \to 0$ as $n \to \infty$, then for each $h \in \{1, 2, ..., n-1\}$ and $\tilde{\rho}(h)$ defined in (2.11), we have

$$\sqrt{n}(\tilde{\rho}(h) - \rho(h)) \Rightarrow N(0, \upsilon_h) \tag{2.13}$$

with the variance $\upsilon_h = \sum_{k=-\infty}^{\infty} (\rho(k+h) + \rho(k-h) - 2\rho(h)\rho(k)))^2$ which is given by Bartlett's formula (Brockwell and Davis, 1991).

Proof: Since

$$a_n(\tilde{\rho}(h) - \rho(h)) = a_n(\tilde{\rho}(h) - \hat{\rho}(h)) + a_n(\hat{\rho}(h) - \rho(h))$$

To show that $a_n(\hat{\rho}(h) - \tilde{\rho}(h)) \to 0$ in probability, we have

$$\begin{aligned} a_n(\hat{\rho}(h) - \tilde{\rho}(h)) &= a_n(\hat{\rho}(h) - w_h\rho_\tau(h) - (1 - w_h)\hat{\rho}(h)) \\ &= a_n w_h(\hat{\rho}(h) - \rho_\tau(h)) \\ &\to 0 \quad \text{in probability,} \end{aligned}$$

the last line follows from the fact that w_h is bounded and the assumption of the theorem. Hence, the proposed estimator of the autocorrelation function has the same asymptotic distribution as the sample ACF. To see Remark 2, we can use the consistency and asymptotic normality of the sample autocorrelation.

Non-negative Definite There is no guarantee that the penalized estimator will provide a non-negative definite solution. However, there are several techniques which can modify the estimated function to be non-negative definite (McMurry and Politis, 2010, 2015). One such technique is described as follows. For $1 \leq m < n$, let $\hat{\mathbf{R}}$ be the $m \times m$ matrix with $\hat{\rho}(i-j)$ in position i, j, and $\tilde{\mathbf{R}}$ be the corresponding symmetric matrix with $\hat{\rho}_\lambda(i-j)$ in position i, j. Both of these matrices are symmetric and real valued, and have a spectral representation with m real valued eigenvalues and m orthogonal eigenvectors vectors each with norm 1. For any symmetric real valued matrix $\mathbf{A}$ and unit norm vector $\mathbf{v}$,

$$\mathbf{v}^t\mathbf{A}\mathbf{v} \geq \gamma_1,$$

where γ_1 is the smallest eigenvalue of $\mathbf{A}$, and the minimum is attained when $\mathbf{v}$ is taken to be the eigenvector corresponding to γ_1. Denote the smallest eigenvalue of $\hat{\mathbf{R}}$ as α and let the corresponding normalized eigenvector be $\hat{\mathbf{e}}$. Assume the estimated variance $\hat{\gamma}(0) > 0$, then $\hat{\mathbf{R}}$ is positive definite and $\alpha > 0$. Now let $\tilde{\mathbf{e}}$ be the normalized eigenvector for $\tilde{\mathbf{R}}$ corresponding to smallest eigenvalue β. If $\beta \geq 0$, then $\tilde{\mathbf{R}}$ is non-negative definite. If $\beta < 0$, we can create a new penalized M-estimator as follows. let

$$\tilde{\rho}_{NND}(h) = c\hat{\rho}(h) + (1 - c)\tilde{\rho}(h),$$

with $c = |\beta|/(\alpha + |\beta|)$ and $h = |i - j|$. This new convex combination of the original penalized M-estimator and the sample ACF is non-negative definite. To see this let $\tilde{\mathbf{R}}_{NND}$ be the $m \times m$ symmetric matrix with i, j element $\tilde{\rho}_{NND}(i-j)$. Without loss of generality, let $m \times 1$ vector $\mathbf{v}$ have unit norm. Using well known properties of quadratic forms in symmetric matrices,

$$\mathbf{v}'\tilde{\mathbf{R}}_{NND}\mathbf{v} \geq c\alpha + (1-c)\beta = 0.$$

Note that the lower bound of zero on the right hand side can only be attained if $\hat{\mathbf{e}} = \pm\tilde{\mathbf{e}}$. Simulations indicate that this procedure only slightly changes the penalized estimator.

Penalized partial correlation estimator The problem of estimating partial correlations can also be formulated in terms of M-estimation. The sample PACF is typically found through use of the Durbin-Levinson algorithm (Brockwell and Davis, 1991). Development similar to that for the ordinary correlation function results in an estimator of the form

$$\tilde{\alpha}(h) = w_h\alpha_\tau(h) + (1 - w_h)\hat{\alpha}(h),$$

where $\hat{\alpha}(h)$ is the ordinary sample PACF, $\alpha_\tau(h)$ is a target value and $w_h = \lambda_h/(1+\lambda_h)$. The target values and tuning parameters can be selected using the sample PACF. Another way to obtain the penalized PACF is decriced in Section 2.4.3 which is calculated from the penalized ACF.

2.3 Selection of Tuning Parameters

In this section, we illustrate how to choose the target and weight using some plug-in approaches and cross-validation (CV).

2.3.1 Plug-in Approaches

Since the bias of the sample correlation increases with the magnitude of correlation, we select the target value based on the sample ACF. We also mimic the idea of adaptive LASSO and allow the penalty λ to depend on the sample ACF as well. The chosen penalties result in a penalized least squares (PLS) estimator which reduces *both* bias and mean squared error relative to the usual estimator.

Partition Method Applying the adaptive lasso and the weights are data-dependent (Zou, 2006), we choose the adaptive weights $\hat{w} = 1/|\hat{\beta}(ols)|$. Similarly, we consider the weight as,

$$\hat{w}_h = \frac{1}{|\hat{\rho}(h)|^2} I(|\hat{\rho}(h)| \leq \ell_h) + \frac{1}{(1-|\hat{\rho}(h)|)^2} I(|\hat{\rho}(h)| > \ell_h)$$

for some $\ell_h > 0$. The resulting solution is given by, $h = 0, 1, ..., n-1$,

$$\tilde{\rho}_{part}(h) = \frac{\hat{\rho}(h) + \lambda_h sgn(\hat{\rho}(h))|\rho_\tau(h)|}{1+\lambda_h} \tag{2.14}$$

where $\lambda_h = \frac{\lambda}{(1-|\hat{\rho}(h)|)^2 \sum_{t=1}^{n-h} Y_t^2} I(|\hat{\rho}(h)| > \ell_h) + \frac{\lambda}{|\hat{\rho}(h)|^2 \sum_{t=1}^{n-h} Y_t^2} I(|\hat{\rho}(h)| \leq \ell_h)$. The thresholds ℓ_h and u_h depend on the series. In practice, we could learn the thresholding rule from data. Based on the asymptotic normality and 95% confidence interval, the choice of $\ell_h = 2/\sqrt{n}$ and $u_h = 1 - 2/\sqrt{n}$ works well for $h = 1$.

We consider $\lambda = f(n)$ which is a function of series size n. We generated a partition of interval $|\rho| \in [0, 1]$ such as, $0 < a_n = \frac{2}{\sqrt{n}} < \frac{2}{\sqrt{n}} + \frac{1}{n} < \frac{2}{\sqrt{n}} + \frac{2}{n} < ... < b_n = 1 - \frac{2}{\sqrt{n}} < 1$. So our target correlations are

$$\rho_\tau(h) = \begin{cases} 0 & |\hat{\rho}(h)| \leq 2/\sqrt{n} \\ \frac{2}{\sqrt{n}} + \frac{1}{n} & 2/\sqrt{n} < |\hat{\rho}(h)| \leq 2/\sqrt{n} + 1/n \\ \vdots & \vdots \\ 1 & |\hat{\rho}(h)| > 1 - 2/\sqrt{n} \end{cases} \tag{2.15}$$

To make the solution continuous, we have

$$\lambda_h = \begin{cases} \frac{(\ell_h - |\hat{\rho}(h)|)}{(n-h)|\hat{\rho}(h)|^2 \overline{Y_t^2}} & |\hat{\rho}(h)| \leq \ell_h \\ \frac{(|\hat{\rho}(h)| - \ell_h)|\rho_\tau(h)|}{(n-h)|1-\hat{\rho}(h)|^2 \overline{Y_t^2}} & |\hat{\rho}(h)| > \ell_h. \end{cases}$$

Bias Correction Method The empirical bias of sample ACF in stationary time series shows that the true correlation with unbiased estimates is less than zero. For example, $\rho(bias = 0) \approx -0.25$ at $h = 1$ and $n = 100$ for AR(1) models. Using this empirical rule, we shrink correlation to zero at $\rho(bias = 0)$ and choose a continuous and soft target function to correct the bias. Let $c = \rho(bias = 0)$,

$\ell_h = 1/\sqrt{n-h}$ and $u_h = 1 - 2.5/\sqrt{n-h}$, the target is

$$\rho_\tau(h) = \begin{cases} (\hat{\rho}(h) + \sqrt{1 - \frac{2}{1+exp\{3(\hat{\rho}(h)-c)\}}})\frac{2+|\hat{\rho}(h)|}{n} & |\hat{\rho}(h)| \leq u_h \\ max\{\hat{\rho}(h) + \frac{1}{n}, 1\} & |\hat{\rho}(h)| > u_h \end{cases} \tag{2.16}$$

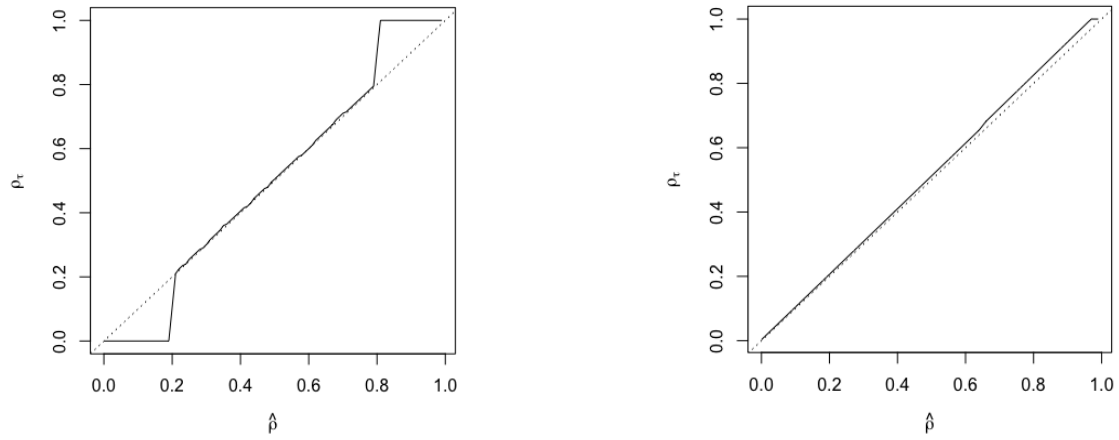
and

$$\lambda_h = \frac{(\ell_h - |\hat{\rho}(h) - c|)^2(1 + sgn(\hat{\rho}(h))\hat{\rho}(h)^2)}{1 - |\hat{\rho}(h)|}$$

then the penalized estimator using bias corrected method is,

$$\tilde{\rho}_{bc}(h) = \frac{\hat{\rho}(h) + \lambda_h sgn(\hat{\rho}(h))|\rho_\tau(h)|}{1 + \lambda_h} \tag{2.17}$$

Figure 2.2 shows the two target functions for partition and bias correction methods. The ρ are restricted between $(0, 1)$ and the target plot is the same for negative correlations.



(a) Partition

(b) Bias Correction

Figure 2.2: (a) ρ_τ for partition estimator (b) ρ_τ for bias correction estimator as functions of $\hat{\rho}(1)$ with $n = 100$, and $h = 1$.

2.3.2 Cross Validation

Bergmeir et al. (2018) introduced K-fold cross validation for dependent data, where the training and test set are the entire rows in the embedded matrix (see Section 1.3). To avoid the underuse of the sample, we use leave-one-out CV, so that each fold consists of only one row of the matrix, and

calculate the penalized estimator based on the forward and backward ACF. The objective function is given in (2.6). Given the time series $y_1, y_2, ..., y_n$ with zero mean, we can set up the training set and test set of the pairs for lag $h = 1, ..., n-1$ as follows, For $t = 1, 2, ..., n$, Select the pair $\{y_t, y_{t+h}\}$ at time t for the test set, and use the pairs at times $1, ..., t-1, t+1, ..., n$ to fit the model. The training set includes $\{y_1, y_{h+1}\}$, $\{y_2, y_{h+2}\}$,... $\{y_{t-1}, y_{t+h-1}\}$, $\{NA, NA\}$,$\{y_{t+1}, y_{t+h+1}\}$,,... $\{y_{n-h}, y_n\}$. Then we can calculate the $\hat{\rho}_{-t}$ using the sample autocorrelation function without the value at time t. For $h < (n-1)/2$,

$$CV_{-t} = \begin{cases} (y_t - \tilde{\rho}_{-t}(h)y_{t+h})^2 & \text{t = 1, ..., h} \\ (y_{t+h} - \tilde{\rho}_{-t}(h)y_t)^2 + (y_t - \tilde{\rho}_{-t}(h)y_{t+h})^2 & \text{t = h+1, ..., n-h} \\ (y_t - \tilde{\rho}_{-t}(h)y_{t-h})^2 & \text{t = n-h+1, ..., n} \end{cases}$$

and for $h \geq (n-1)/2$,

$$CV_{-t} = \begin{cases} (y_t - \tilde{\rho}_{-t}(h)y_{t+h})^2 & \text{t = 1, ..., n-h} \\ 0 & \text{t = n-h+1, ..., h} \\ (y_t - \tilde{\rho}_{-t}(h)y_{t-h})^2 & \text{t = h+1, ..., n} \end{cases}$$

Then the CV calculated by

$$CV = \frac{1}{2(n-h)} \sum_{t=1}^{n} CV_{-t}$$

For the CV method, we use the bias correction target in (2.16) and let w_h range from 0 to 1 by 0.001 for all cases. Let the selected weight be w_h^*, the penalized estimator using leave one out cross validation technique is

$$\tilde{\rho}_{cv}(h) = w_h^* \rho_\tau(h) + (1 - w_h^*)\hat{\rho}(h) \tag{2.18}$$

The limitation of leave-one-out cross-validation can be very time consuming to implement in general, but it is fast to compute CV for linear models.

2.4 Simulations

In order to examine the finite performance of the proposed estimator, a simulation study is conducted to generate stationary time series from different ARMA models. In this section, the proposed estimator $\tilde{\rho}$ is compared with the sample ACF $\hat{\rho}$ in three ways: bias, the ratio of two mean squared errors (MSE) and the distribution of the ratios of two squared errors in one replicate.

- **Empirical Bias:** the absolute empirical bias is calculated by, for $h = 1, 2, ..., n-1$

$$Bias_{\hat{\rho}} = |\overline{\hat{\rho}(h)} - \rho(h)|$$

$$Bias_{\tilde{\rho}} = |\overline{\tilde{\rho}(h)} - \rho(h)|$$

- **Ratio of two mean squared errors (RMSE):** the ratio of two MSEs is to compare the performance of proposed estimator and sample autocorrelation estimator. Let $MSE_{\hat{\rho}}$ be the mean squared error of the conventional estimator and $MSE_{\tilde{\rho}}$ be the mean squared error of the proposed estimator, we define RMSE [1] to be the ratio of $MSE_{\hat{\rho}}$ and $MSE_{\tilde{\rho}}$ as follows, assume each simulation is repeated M times,

$$MSE_{\hat{\rho}} = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{h}\left(\hat{\rho}(i)^{(m)} - \rho(i)\right)^2$$

$$MSE_{\tilde{\rho}} = \frac{1}{M}\sum_{m=1}^{M}\sum_{i=1}^{h}\left(\tilde{\rho}(i)^{(m)} - \rho(i)\right)^2$$

$$RMSE = \frac{MSE_{\hat{\rho}}}{MSE_{\tilde{\rho}}}$$

  Both $MSE_{\hat{\rho}}$ and $MSE_{\tilde{\rho}}$ can range from 0 to ∞ and are negatively-oriented scores, which means the lower MSE values the better. In other words, if $RMSE$ values are larger than one, it means that the proposed estimator $\tilde{\rho}$ is better than $\hat{\rho}$.

- **Ratios of two squared errors (RSE):** we are also interested in the distribution of the ratios of two squared errors (SE)). Let $SE_{\hat{\rho}}^{(m)}$ be the squared error of the sample estimator in m-th replicate and $SE_{\tilde{\rho}}^{(m)}$ be the squared error of the proposed estimator, we define RSE to be the ratios of $SE_{\hat{\rho}}^{(m)}$ with respect to $SE_{\tilde{\rho}}^{(m)}$ in m-th replicate as follows, for $m = 1, ..., M$

[1] the RMSE here is not the root mean squared error.

and $h = 1, 2, ..., n-1$

$$SE_{\hat{\rho}}^{(m)} = \sum_{i=1}^{h} \left(\hat{\rho}(i)^{(m)} - \rho(i) \right)^2$$

$$SE_{\tilde{\rho}}^{(m)} = \sum_{i=1}^{h} \left(\tilde{\rho}(i)^{(m)} - \rho(i) \right)^2$$

$$RSE^{(m)} = \frac{SE_{\hat{\rho}}^{(m)}}{SE_{\tilde{\rho}}^{(m)}}$$

2.4.1 Study on ARMA Models with Single Lags

This section presents a variety of simulation experiments. Typically, to model the correlation structure of a time series, we use a stationary and invertible ARMA process of the form

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$$

where p is the AR order, q is the MA order, and $\{\epsilon_t\}$ is an iid innovation sequence of zero mean random variables with finite variance. In this section, we conduct simulation study using various time series models with length $n = 50, 100, 500$. Three different simulated examples are considered as follows,

AR(1): $X_t = \phi X_{t-1} + \epsilon_t$. $\phi = -0.99, -0.95, -0.90, -0.70, -0.50, -0.30, -0.10, 0.10, 0.30, 0.50, 0.70, 0.90, 0.95, 0.99$.

AR(2): $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t$. $\phi_1 = 0$ and $\phi_2 = -0.99, -0.95, -0.90, -0.70, -0.10, 0.10, 0.70, 0.90, 0.95, 0.99$.

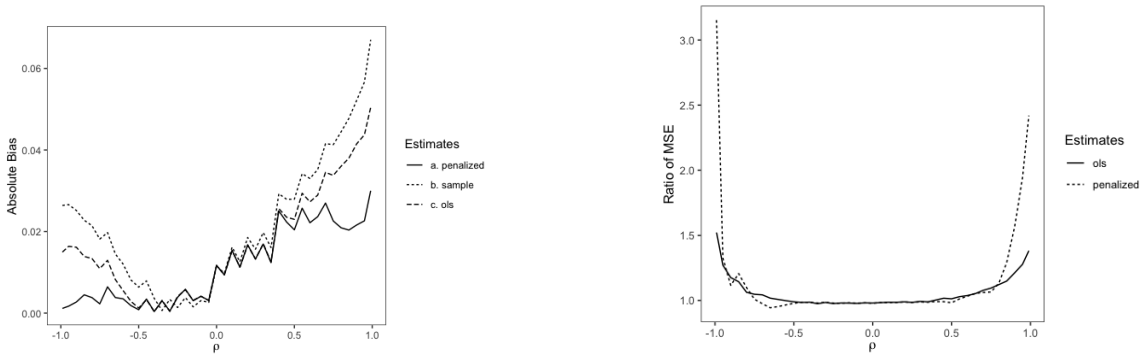
MA(1): $X_t = \theta \epsilon_{t-1} + \epsilon_t$. $\rho(1) = \theta/(1+\theta^2)$ to get $\rho(1) = -0.50, -0.4, -0.10, 0.10, 0.4, 0.50$.

ARMA(1,1): $X_t = \phi X_{t-1} + \theta \epsilon_{t-1} + \epsilon_t$. $\theta = \pm 1$ and $\phi = 2\rho(1) \mp 1$ to get $\rho_(1) = -0.99, -0.95, -0.90, -0.70, -0.10, 0.10, 0.70, 0.90, 0.95, 0.99$.

Throughout this section, we assume $\{\epsilon_t\}$ follows the standard normal distribution. Each simulation was repeated 1000 times. Note the autocorrelation of AR(p), MA(q) and ARMA(p,q) are reviewed in Appendix A.

Table 2.2 shows the the empirical bias, RMSE and RSE for AR(1) model at lag one. Three penalized ACF estimators $\tilde{\rho}_{part}$, $\tilde{\rho}_{bc}$ and $\tilde{\rho}_{cv}$ are compared with the sample ACF. All three proposed

estimators alleviates the underestimation of sample acf for all correlation cases and different sample sizes. When the correlation is strong, all three penalized estimators attain smaller mean squared error compared with the sample ACF, especially the penalized estimator via cross validation. For example, the ratio of MSE of $\hat{\rho}$ and $\tilde{\rho}_{cv}$ is 1.906 when $\rho(1) = 0.95$ and $n = 50$. Last three columns in Table 2.2 show the proportion that $RSE > 1$ in 1000 replicates. For strong and moderate correlations, over 50% times that penalized estimators performs better than the sample ACF. Also, the proposed estimators are competitive with the sample ACF in mean squared error. The estimator $\tilde{\rho}_{part}$ outperforms the sample ACF when the correlation is weak for all $\rho(1)$ and n cases. Compare to the penalized estimators using partition method and cross validation, the estimator using bias correction method is more consistent, even though its ratio of MSEs are not as good as the ones for $\tilde{\rho}_{part}$ and $\tilde{\rho}_{cv}$. The worst case in RMSE for $\tilde{\rho}_{bc}$ is 0.980 for $n = 100$ which is very close to sample ACF. For lag one case, we are also interested in comparing the proposed estimator with other alternative estimators other than sample ACF. Although some estimators such as C statistic, Translated ACF estimator performs better than sample ACF with moderate positive correlation, these performs worse for all negative correlations compared with the sample ACF and penalized ACF estimators in terms of bias and MSE. Figure 2.3 shows the comparison between least square estimator (ols) and penalized estimator using CV in absolute bias when $n = 100$ and $h = 1$ for AR(1) model. We can see that the proposed estimator generally outperforms the sample ACF and least square estimator in terms of both bias and MSE.



(a) Bias

(b) RMSE

Figure 2.3: (a) Absolute bias of penalized, sample and least squares estimates (b) Ratio of mean squared errors of penalized and least squares estimates with respect to sample ACF under AR(1) model, $n = 100$, and $h = 1$.

Table 2.2: Absolute Bias (Bias), ratio of mean squared errors (RMSE) and ratio of squared errors (RSE) for AR(1) models, $n = 50, 100, 500$, $h = 1$

		Bias				RMSE			$Pr\{RSE > 1\}$		
n	$\rho(1)$	$\hat{\rho}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$
50	-0.99	0.047	0.013	0.033	0.006	2.323	1.345	**2.983**	0.860	0.989	0.879
	-0.95	0.051	0.013	0.036	0.005	**1.702**	1.136	1.614	0.624	0.694	0.556
	-0.90	0.050	0.012	0.037	0.009	**1.322**	1.033	1.254	0.587	0.664	0.598
	-0.70	0.029	0.002	0.025	0.003	0.924	**0.979**	0.920	0.498	0.544	0.498
	-0.50	0.019	0.005	0.019	0.008	0.916	**0.990**	0.959	0.469	0.508	0.504
	-0.30	0.003	0.001	0.003	0.003	0.873	**0.998**	0.965	0.170	0.271	0.476
	-0.10	0.019	0.013	0.019	0.021	**1.009**	0.999	0.958	0.513	0.285	0.250
	0.10	0.025	0.028	0.024	0.023	**1.048**	0.988	0.963	0.567	0.537	0.246
	0.30	0.049	0.049	0.046	0.044	0.897	0.978	0.987	0.204	0.625	0.587
	0.50	0.064	0.052	0.054	0.052	0.975	0.969	**1.007**	0.589	0.662	0.661
	0.70	0.081	0.058	0.055	0.052	**1.094**	1.068	1.059	0.717	0.710	0.693
	0.90	0.107	0.065	0.063	0.050	1.459	1.384	**1.589**	0.820	0.801	0.772
	0.95	0.119	0.074	0.072	0.055	1.560	1.527	**1.906**	0.921	0.894	0.884
	0.99	0.135	0.086	0.086	0.065	1.713	1.668	**2.288**	1.000	1.000	1.000
100	-0.99	0.026	0.008	0.016	0.002	2.130	1.392	**2.632**	0.746	0.839	0.685
	-0.95	0.028	0.006	0.019	0.001	**1.527**	1.150	1.372	0.613	0.675	0.554
	-0.90	0.024	0.004	0.017	0.002	**1.253**	1.033	1.115	0.557	0.612	0.557
	-0.70	0.014	0.001	0.012	0.000	**1.003**	0.986	0.955	0.489	0.529	0.490
	-0.50	0.011	0.003	0.011	0.006	0.979	**0.993**	**0.993**	0.525	0.536	0.529
	-0.30	0.002	0.006	0.002	0.005	0.945	**0.999**	0.980	0.385	0.336	0.473
	-0.10	0.006	0.003	0.006	0.007	0.978	**0.999**	0.979	0.381	0.350	0.332
	0.10	0.011	0.014	0.010	0.010	**1.000**	0.994	0.982	0.432	0.521	0.334
	0.30	0.019	0.017	0.018	0.016	0.954	0.984	**0.986**	0.428	0.561	0.554
	0.50	0.025	0.017	0.019	0.017	**0.994**	0.980	0.978	0.582	0.590	0.587
	0.70	0.039	0.027	0.025	0.025	**1.070**	1.026	1.053	0.636	0.627	0.640
	0.90	0.053	0.031	0.028	0.022	1.389	1.376	1.554	0.745	0.723	0.704
	0.95	0.057	0.031	0.030	0.022	1.643	1.574	**1.998**	0.831	0.810	0.794
	0.99	0.069	0.041	0.041	0.030	1.812	1.747	**2.441**	0.996	0.996	0.999
500	-0.99	0.006	0.001	0.003	0.001	1.692	1.261	**1.771**	0.559	0.633	0.509
	-0.95	0.005	0.000	0.003	0.001	1.119	1.063	**1.133**	0.516	0.552	0.493
	-0.90	0.005	0.000	0.003	0.001	**1.049**	1.022	1.018	0.540	0.559	0.521
	-0.70	0.003	0.000	0.002	0.001	**0.996**	**0.996**	0.983	0.509	0.522	0.507
	-0.50	0.002	0.001	0.002	0.001	0.998	0.998	**1.002**	0.508	0.518	0.512
	-0.30	0.003	0.004	0.003	0.004	0.992	**1.000**	0.996	0.460	0.455	0.464
	-0.10	0.001	0.001	0.001	0.001	0.972	**1.000**	0.996	0.163	0.498	0.479
	0.10	0.001	0.001	0.001	0.001	0.968	**0.998**	0.996	0.157	0.506	0.490
	0.30	0.005	0.004	0.005	0.005	0.997	0.997	**1.000**	0.537	0.540	0.539
	0.50	0.005	0.003	0.004	0.003	**1.000**	0.996	0.989	0.520	0.524	0.521
	0.70	0.007	0.005	0.005	0.003	1.016	1.014	**1.021**	0.570	0.569	0.563
	0.90	0.010	0.005	0.004	0.003	**1.148**	1.132	1.138	0.602	0.600	0.589
	0.95	0.010	0.005	0.004	0.003	1.288	1.294	**1.395**	0.654	0.645	0.631
	0.99	0.013	0.007	0.007	0.005	1.714	1.690	**2.119**	0.864	0.832	0.810

Table 2.3: Absolute Bias (Bias), ratio of mean squared errors (RMSE) for ARMA(1,1) and MA(1) models, $n = 50, 100, 500$, $h = 1$

			Bias				RMSE		
Model	n	$\rho(1)$	$\hat{\rho}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$
ARMA(1,1)	50	-0.99	0.038	0.009	0.024	0.002	2.390	1.601	**5.241**
		-0.9	0.032	0.003	0.018	0.003	**1.332**	1.037	1.193
		-0.7	0.029	0.007	0.026	0.007	**1.031**	0.986	0.974
		-0.1	0.002	0.007	0.002	0.000	**1.025**	0.998	0.960
		0.1	0.021	0.025	0.021	0.020	**1.051**	0.989	0.963
		0.7	0.052	0.032	0.023	0.023	**1.100**	1.069	1.032
		0.9	0.061	0.028	0.013	0.003	1.615	1.639	**1.924**
		0.99	0.084	0.041	0.033	0.015	2.283	2.727	**6.347**
	100	-0.99	0.021	0.005	0.011	0.004	2.395	1.704	**4.582**
		-0.9	0.020	0.004	0.012	0.000	**1.275**	1.070	1.077
		-0.7	0.016	0.006	0.014	0.004	**1.035**	0.994	0.979
		-0.1	0.000	0.004	0.000	0.001	0.978	**0.999**	0.981
		0.1	0.013	0.017	0.013	0.012	0.972	**0.995**	0.982
		0.7	0.021	0.010	0.006	0.008	**1.065**	0.987	1.026
		0.9	0.031	0.014	0.005	0.001	1.374	1.423	**1.567**
		0.99	0.041	0.019	0.013	0.005	2.224	2.814	**5.278**
	500	-0.99	0.004	0.004	0.003	0.003	0.992	**0.999**	0.997
		-0.9	0.003	0.002	0.002	0.002	0.990	**0.998**	0.997
		-0.7	0.004	0.003	0.003	0.003	0.994	0.996	**0.998**
		-0.1	0.006	0.002	0.000	0.001	1.273	1.232	**1.334**
		0.1	0.002	0.002	0.002	0.002	0.967	**0.999**	0.997
		0.7	0.005	0.003	0.002	0.001	1.015	1.016	**1.017**
		0.9	0.005	0.002	0.000	0.001	1.106	1.110	**1.115**
		0.99	0.007	0.003	0.001	0.000	1.949	2.045	**2.841**
MA(1)	50	-0.5	0.020	0.007	0.005	0.009	0.959	**0.990**	0.985
		-0.4	0.010	0.002	0.005	0.003	0.929	**0.996**	0.982
		-0.1	0.015	0.010	0.025	0.018	**1.004**	0.998	0.957
		0.1	0.024	0.029	0.017	0.023	**1.039**	0.990	0.967
		0.4	0.038	0.030	0.023	0.030	0.935	0.960	**1.004**
		0.5	0.049	0.038	0.034	0.038	1.005	0.961	**1.009**
	100	-0.5	0.005	0.001	0.003	0.000	0.981	0.993	**1.000**
		-0.4	0.005	0.000	0.003	0.001	0.966	**0.997**	0.986
		-0.1	0.007	0.004	0.013	0.008	0.984	**0.999**	0.979
		0.1	0.014	0.017	0.008	0.013	**1.007**	0.995	0.983
		0.4	0.020	0.015	0.012	0.016	0.990	0.983	**0.996**
		0.5	0.019	0.012	0.011	0.012	**1.017**	0.983	0.982
	500	-0.5	0.004	0.002	0.002	0.003	1.000	0.999	1.004
		-0.4	0.000	0.001	0.002	0.001	0.992	**0.999**	0.995
		-0.1	0.002	0.001	0.003	0.002	0.968	**1.000**	0.996
		0.1	0.002	0.003	0.001	0.002	0.963	**0.999**	0.997
		0.4	0.006	0.005	0.004	0.005	1.003	0.999	**1.005**
		0.5	0.005	0.003	0.003	0.003	**1.007**	1.001	0.997

Table 2.4: Absolute Bias (Bias), ratio of mean squared errors (RMSE) for AR(2) models, $n = 50, 100, 500$, $h = 2$ and $\phi_1 = 0$

			Bias				RMSE		
Model	n	ϕ_2	$\hat{\rho}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$
AR(2)	50	-0.99	0.055	0.006	0.037	0.002	4.137	1.581	**8.711**
		-0.9	0.062	0.006	0.050	0.002	1.651	1.057	**1.622**
		-0.7	0.051	0.011	0.048	0.011	**1.066**	0.992	1.053
		-0.1	0.011	0.009	0.011	0.015	0.968	**0.998**	0.918
		0.1	0.032	0.035	0.031	0.029	**1.011**	0.991	0.932
		0.7	0.090	0.052	0.065	0.045	**1.201**	1.072	1.158
		0.9	0.115	0.054	0.072	0.039	1.799	1.405	**2.027**
		0.99	0.092	0.034	0.042	0.021	2.387	1.877	**3.597**
	100	-0.99	0.032	0.004	0.021	0.002	3.262	1.508	**5.492**
		-0.9	0.035	0.005	0.028	0.004	**1.442**	1.052	1.352
		-0.7	0.022	0.002	0.020	0.001	**1.017**	0.992	0.987
		-0.1	0.005	0.002	0.005	0.007	0.956	**0.999**	0.957
		0.1	0.017	0.020	0.017	0.016	0.979	**0.995**	0.967
		0.7	0.046	0.025	0.032	0.024	1.140	1.034	**1.121**
		0.9	0.058	0.025	0.033	0.016	1.659	1.406	**1.895**
		0.99	0.052	0.020	0.024	0.011	2.450	1.961	**4.013**
	500	-0.99	0.007	0.000	0.005	0.002	2.160	1.343	**2.337**
		-0.9	0.007	0.000	0.005	0.001	**1.102**	1.033	1.074
		-0.7	0.004	0.000	0.003	0.001	**1.008**	0.998	0.998
		-0.1	0.000	0.000	0.000	0.001	0.966	**1.000**	0.992
		0.1	0.004	0.004	0.004	0.003	0.959	**0.999**	0.993
		0.7	0.009	0.004	0.006	0.003	1.030	1.019	**1.035**
		0.9	0.012	0.005	0.006	0.003	**1.240**	1.156	1.232
		0.99	0.012	0.005	0.006	0.003	2.140	1.784	**2.989**

Table 2.3 and 2.4 display the empirical bias and ratio of the MSE for ARMA(1,1), MA(1) and AR(2) at $h = 1$ or $h = 2$. For each parameter setting, sample size is taken as $n = 50, 10, 500$. We compare the performance between the proposed penalized estimator and sample ACF at lag one for ARMA(1,1) and MA(1), and at lag two for AR(2) model. Tables provide a summary of the estimates from weak to high correlation with different sample size for ARMA(1,1) and MA(1). According to the relationship between $\rho(1)$ and θ, we consider taking the parameters $\rho = 0.1, ..., 0.5$. From these results, one will see that the proposed estimators reduce the bias for all cases and generally have

smaller MSE when $\rho \geq 0.7$. The penalized estimator using partition method has better performance for most of the ARMA(1,1) cases, especially when $n = 50$. Similarly, the penalized estimator via CV performs the best in RMSE when $\rho \geq 0.9$. Overall, the results show that the proposed estimators outperform the sample ACF in most cases, and the competitive performance for strongly correlated time series at single lag.

2.4.2 Study on ARMA Models with Multiple Lags

Next, we simulate time series of length 100 and discuss the performance of the penalized autocorrelation at multiple lags $\{1, 2, ..., p_n\}$ where $p_n = (n)^{1/2} = 10$ (Bickel and Gel, 2011). Four different stationary time series models are generated as follows,

$$\text{AR}(1) : X_t = \phi_1 X_{t-1} + \epsilon_t; \phi_1 = -0.99, -0.9, -0.5, -0.1, 0.1, 0.5, 0.9, 0.99.$$

$$\text{AR}(2) : X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t; \phi_1 = 0, \phi_2 = -0.99, -0.9, -0.5, -0.1, 0.1, 0.5, 0.9, 0.99.$$

$$\text{MA}(1) : X_t = \theta_1 \epsilon_{t-1} + \epsilon_t; \theta_1 = -0.99, -0.9, -0.5, -0.1, 0.1, 0.5, 0.9, 0.99.$$

$$\text{MA}(2) : X_t = \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \epsilon_t; \theta_1 = 0, \theta_2 = -0.99, -0.9, -0.5, -0.1, 0.1, 0.5, 0.9, 0.99.$$

In order to obtain better estimation for multiple lags cases, we construct a band and taper estimates for penalized ACF and then compare with the sample ACF. McMurry and Politis (2010) proposed a banded and tapered estimates for autocovariance matrices, which gradually down-weights the correlation to zero starting from l. The detail of selecting the banding parameter l see McMurry and Politis (2015). Generally, the weights can given by

$$\kappa(x) = \begin{cases} 1 & \text{if } |x| \leq l \\ g(|x|) & \text{if } l < |x| \leq 2l \\ 0 & \text{if } |x| > 2l. \end{cases}$$

For the estimation of ACF in multiple lags when $|\rho(h)| < 0.7$, we consider use the penalized estimators when $h \leq l$, sample ACF when $l < h \leq max\{2l, p_n\}$ and 0 for $h > max\{2l, p_n\}$. Table 2.5 reports the RMSE for each above process. Note that without using the banded and tapered method, the RMSE of the estimates with multi lags are close to or larger than one for all AR and MA models. There is the comparatively larger bias of the proposed estimator between ℓ_h and u_h

since the ρ might be shrunk to the opposite direction near thresholds. So it is evident that the ratio of MSE for moderate correlation is less than 1. For MA(1) models, the performance of the proposed estimator is better if the serial correlation is weak. Based on the autocorrelation of MA(1) model in Appendix A, we have $|\rho(1)| < 0.5$ when $\theta_1 = 0.99$. The RMSE of the proposed estimator is less than one but very close to one. In addition, the proposed estimator has almost the same performance as the conventional estimator for AR(2) and MA(2) cases in the simulation. After applying the banded and tapered method, the results in Table 2.5 show that the penalized ACF estimator with banded and tapered weights performs much better than the sample ACF with respect to MSE. To sum up, this simulation study shows that the proposed estimator is competitive and even much better than the sample ACF in terms of RMSE when there are multiple lags considered for estimation of the ACF.

Table 2.5: Ratio of mean squared errors (RMSE) of the penalized autocorrelation estimators for AR and MA cases with multi-lag, $n = 100$, $h = 10$

			RMSE(ρ)						RMSE(ρ)		
Model	ϕ_1	ϕ_2	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$	Model	θ_1	θ_2	$\tilde{\rho}_{part}$	$\tilde{\rho}_{bc}$	$\tilde{\rho}_{cv}$
AR(1)	-0.99	-	1.589	1.210	1.977	MA(1)	-0.99	-	3.922	3.973	6.221
	-0.9	-	1.101	1.000	1.091		-0.9	-	4.205	4.247	7.650
	-0.5	-	2.241	2.283	1.532		-0.5	-	4.078	4.123	7.043
	-0.1	-	4.465	4.519	7.959		-0.1	-	4.192	4.239	7.411
	0.1	-	4.399	4.401	7.978		0.1	-	4.635	4.619	8.899
	0.5	-	2.352	2.366	1.540		0.5	-	4.314	4.359	7.603
	0.9	-	1.153	1.054	1.173		0.9	-	3.803	3.833	6.163
	0.99	-	1.365	1.154	1.535		0.99	-	4.002	4.039	6.794
AR(2)	-	-0.99	2.329	1.534	3.164	MA(2)	-	-0.99	2.506	2.491	3.357
	-	-0.9	1.292	1.071	1.239		-	-0.9	2.418	2.409	3.240
	-	-0.5	1.630	1.636	1.111		-	-0.5	2.871	2.834	2.694
	-	-0.1	5.244	4.992	4.550		-	-0.1	4.996	4.773	4.447
	-	0.1	4.045	3.922	3.818		-	0.1	4.114	3.978	3.894
	-	0.5	1.505	1.516	1.114		-	0.5	2.301	2.289	2.104
	-	0.9	0.989	1.010	0.986		-	0.9	2.196	2.194	2.708
	-	0.99	0.963	0.995	0.945		-	0.99	2.047	2.046	2.413

2.4.3 Penalized Partial Correlation Estimator

Recall that the sample partial autocorrelation is defined as $\hat{\alpha}(h) = \hat{\boldsymbol{R}}_h^{-1}\boldsymbol{\rho}_h$, where $\hat{\boldsymbol{R}}_h = [\hat{\rho}(i-j)]_{i,j=1}^{h}$, and $\hat{\boldsymbol{\rho}}_h = (\hat{\rho}(1), ..., \hat{\rho}(h))'$ in Section 3.1. We also can obtain the penalized PACF as

$$\tilde{\alpha}(h) = \tilde{\boldsymbol{R}}_h^{-1}\tilde{\boldsymbol{\rho}}_h,$$

where $\boldsymbol{R}_h = [\tilde{\rho}(i-j)]_{i,j=1}^{h}$, and $\tilde{\boldsymbol{\rho}}_h = (\hat{\rho}(1), ..., \hat{\rho}(h))'$. This section presents a simulation study under AR(2) models to examine the performance of penalized partial autocorrelation against the usual sample PACF.

Table 2.6 provides the ratio of MSE for penalized PACF estimators under AR(2) models. In each case, 1000 replicates were performed where an AR(2) process with size $n = 50, 100, 500$ was generated using several ϕ_1 and ϕ_2 values. Note that in AR(2), $\alpha(1) = \phi_1/(1-\phi_2)$ and $\alpha(2) = \phi_2$. Parameters ϕ_1 and ϕ_2 are taken with different combination to have $\alpha(1) = 0.5, 0.7, 0.9$ and $\alpha(2) = 0.6, 0.7, 0.8, 0.9$. We see in Table 2.6 the RMSE of proposed PACF estimators are better than or close to the sample PACF for all cases at lag 1. The penalized PACF estimators $\tilde{\alpha}(2)$ have smaller absolute bias and MSE compare with the sample PACF $\hat{\alpha}(2)$ for almost all parameters and the different sample sizes considered.

2.5 Discussion

This chapter introduces new estimation of (partial) autocorrelation function for stationary time series, which can alleviate the bias and reduce the mean squared error of the usual sample ACF. Using the penalized methods, the proposed estimators shrink towards the target. The resulting penalized estimators can achieve the sample acf, least square and forward and backward correlation estimator as a special case. The proposed ACF/PACF estimators are easy to compute, but still tend to perform better in estimation than the usual sample ACF/PACF. In many cases, especially when the correlation is strong, the penalized estimators have smaller bias and MSE compared with those in the literature. The simulations considered in this chapter demonstrate that the penalized scheme can improve estimation of (partial) autocorrelation for ARMA processes.

Table 2.6: Empirical Bias (Bias) and ratio of mean squared errors (RMSE) of the penalized partial autocorrelation estimators under AR(2) models, $\alpha_1 = \alpha(1)$, $\alpha_2 = \alpha(2)$.

			RMSE($\alpha(1)$)			Bias($\alpha(2)$)				RMSE($\alpha(2)$)		
n	α_1	α_2	$\tilde{\alpha}_1^{part}$	$\tilde{\alpha}_1^{bc}$	$\tilde{\alpha}_1^{cv}$	$\hat{\alpha}_2$	$\tilde{\alpha}_2^{part}$	$\tilde{\alpha}_2^{bc}$	$\tilde{\alpha}_2^{cv}$	$\tilde{\alpha}_2^{part}$	$\tilde{\alpha}_2^{bc}$	$\tilde{\alpha}_2^{cv}$
50	0.5	0.6	0.989	0.988	0.987	0.133	0.096	0.115	0.094	1.170	1.032	1.155
	0.7	0.6	1.014	1.014	1.030	0.135	0.085	0.116	0.085	1.187	0.998	1.192
	0.9	0.6	1.066	1.062	1.102	0.165	0.097	0.157	0.103	1.116	0.834	1.126
	0.5	0.7	0.991	0.992	0.985	0.143	0.094	0.115	0.089	1.303	1.122	1.316
	0.7	0.7	1.014	1.009	1.015	0.160	0.102	0.132	0.096	1.334	1.097	1.374
	0.9	0.7	1.044	1.033	1.059	0.182	0.097	0.157	0.098	1.405	1.037	1.448
	0.5	0.8	0.987	0.996	0.982	0.167	0.106	0.131	0.095	1.470	1.243	1.573
	0.7	0.8	1.003	1.005	1.006	0.169	0.100	0.132	0.090	1.540	1.232	1.633
	0.9	0.8	1.017	1.015	1.024	0.194	0.104	0.157	0.098	1.712	1.189	1.866
	0.5	0.9	0.981	0.997	0.974	0.174	0.099	0.125	0.080	1.806	1.447	2.160
	0.7	0.9	0.991	1.001	0.988	0.184	0.106	0.137	0.088	1.818	1.410	2.179
	0.9	0.9	1.000	1.005	1.000	0.186	0.103	0.139	0.088	1.929	1.425	2.273
100	0.5	0.6	0.990	0.987	0.991	0.058	0.036	0.045	0.035	1.134	1.039	1.127
	0.7	0.6	1.034	1.012	1.031	0.070	0.040	0.054	0.038	1.180	1.040	1.202
	0.9	0.6	1.089	1.083	1.127	0.097	0.039	0.085	0.041	1.276	0.837	1.534
	0.5	0.7	0.993	0.992	0.995	0.066	0.038	0.048	0.034	1.248	1.124	1.280
	0.7	0.7	1.024	1.011	1.026	0.076	0.039	0.053	0.033	1.323	1.188	1.428
	0.9	0.7	1.062	1.054	1.092	0.105	0.038	0.079	0.035	1.475	1.124	1.835
	0.5	0.8	0.996	0.997	0.995	0.080	0.047	0.056	0.039	1.393	1.256	1.496
	0.7	0.8	1.009	1.005	1.012	0.087	0.043	0.059	0.035	1.504	1.316	1.728
	0.9	0.8	1.030	1.026	1.048	0.106	0.041	0.073	0.033	1.799	1.396	2.385
	0.5	0.9	0.995	0.998	0.993	0.089	0.047	0.060	0.038	1.794	1.469	2.188
	0.7	0.9	1.000	1.002	1.001	0.098	0.053	0.068	0.043	1.934	1.533	2.354
	0.9	0.9	1.008	1.007	1.013	0.105	0.050	0.072	0.038	2.024	1.561	2.697
500	0.5	0.6	1.005	0.999	0.994	0.012	0.007	0.010	0.006	1.050	1.025	1.069
	0.7	0.6	1.015	1.012	1.026	0.013	0.006	0.010	0.005	1.074	1.023	1.091
	0.9	0.6	1.077	1.073	1.113	0.021	0.002	0.013	0.003	1.138	1.053	1.242
	0.5	0.7	1.003	0.998	0.992	0.011	0.005	0.008	0.004	1.061	1.039	1.096
	0.7	0.7	1.016	1.010	1.024	0.016	0.008	0.011	0.007	1.168	1.072	1.186
	0.9	0.7	1.052	1.053	1.095	0.023	0.001	0.010	0.001	1.241	1.191	1.472
	0.5	0.8	1.001	0.998	0.995	0.014	0.007	0.009	0.006	1.182	1.079	1.216
	0.7	0.8	1.011	1.008	1.016	0.016	0.005	0.008	0.004	1.261	1.178	1.286
	0.9	0.8	1.032	1.035	1.065	0.026	0.001	0.010	0.002	1.537	1.468	1.924
	0.5	0.9	1.001	0.999	0.998	0.017	0.008	0.010	0.006	1.404	1.291	1.506
	0.7	0.9	1.007	1.005	1.010	0.019	0.007	0.010	0.005	1.561	1.436	1.799
	0.9	0.9	1.017	1.015	1.031	0.028	0.004	0.013	0.001	1.786	1.711	2.666

A current challenge is to find the optimal λ under the given target, so that we can reduce the

MSE for general ARMA process and all correlation parameter values. There is a trade-off between reducing MSE under strong correlation $\rho \to 1$ or $\rho \to 0$, and keeping the performance the same as sample ACF with moderate correlation. The proposed estimator will underestimate the correlation if the shrinkage moves in the wrong direction. We suggest the estimation of (partial) autocorrelation function for moderately correlated time series as our future work.

Chapter 3

Penalized ACF/PACF for Time Series Goodness of Fit

3.1 Introduction

Much of modeling of stationary time series is based on estimates of the sample autocorrelation function (ACF) and partial autocorrelation function (PACF). In this chapter, our focus is on time series goodness of fit procedures, which are based on estimates of the sample ACF and PACF of the residuals from a fitted model (Ljung and Box, 1978; Li and McLeod, 1981; Monti, 1994; Li and Mak, 1994; Peña and Rodríguez, 2002; Mahdi and McLeod, 2012; Fisher and Gallagher, 2012) . The procedures mentioned above are dependent on accurate estimates of the magnitude of correlation, which when underestimated results in a loss of power in detecting model under fit. However, the traditional sample ACF and PACF tend to underestimate the magnitude of the correlation, and that the bias of these estimators increases with the magnitude of the correlation (Marriot and Pope, 1954). The general bias formula for the sample ACF depends on the entire correlation function (e.g., see Dürre et al. (2015)) and model specific bias corrected estimates while reducing the bias, typically have no impact on, or can even increase MSE (MacKinnon and Smith Jr, 1998; Reschenhofer, 2019). In the first order autoregressive case with heteroscedastic errors and potential unit roots median unbiased estimators (Andrews, 1993; Reschenhofer, 2019) can improve over the sample ACF, but these estimators have higher MSE in the simple stationary first order autoregressive model with

independent Gaussian innovations. In this chapter we consider improving the estimation of (partial) autocorrelation for stationary time series via penalized object function minimization. The resulting estimators can be used to perform a variety of time series modeling procedures, but our focus here will be on improving the performance of time series goodness of fit procedures.

Several authors have considered shrinking (partial) correlations toward zero. A recent sequence of papers have considered banded and tapered estimators (Bickel and Gel, 2011; McMurry and Politis, 2015; Proietti and Giovannelli, 2018), which are equivalent to taking $\rho_\tau = 0$ in (1.12), and letting λ be a function only of the lag; in these papers, small lag (partial) correlations are not shrunk, moderate lag (partial) correlations are shrunk toward zero, and large lag (partial) correlations are set to 0. Alternatively, Liao et al. (2016) shrink correlations toward zero and allow their tuning parameter to depend on the data, as opposed to simply the lag under consideration. Since these methods shrink estimated correlations toward zero, they cannot increase power of goodness of fit tests.

In this chapter we use ordinary and partial correlation estimators found via minimizing (1.12) to increase power of time series goodness of fit for linear and nonlinear models, and tests for detecting nonlinear effects. Judicious choice of target and tuning parameter in (1.12) can shrink small magnitude correlations toward zero to control type I error, and move squared moderate magnitude correlations toward unity to increase power. Careful data based choices of ρ_τ and λ can ensure smooth translation from non-penalized estimator $\hat{\rho}$ to penalized solution $\tilde{\rho}_\lambda$, while retaining the same asymptotic behavior as $\hat{\rho}$. Specific data based choices are provided for general classes of time series Portmanteau tests, these classes include the most prevalent tests appearing in the time series goodness of fit literature. The methods proposed below are novel in two distinct ways. First, the target value is chosen based on an initial estimate of correlation and can be larger in absolute value than $\hat{\rho}$, so that the direction of movement can be away from 0. Second, the target value and tuning parameter are chosen to depend on the nominal type I error bound of a hypothesis test, which does not impact the type I error properties of the test, but can increase power. Simulations show that using $\tilde{\rho}_\lambda$ instead of the sample ACF $\hat{\rho}$, and using a penalized partial correlation $\tilde{\pi}_\lambda$ instead of the PACF $\hat{\pi}$ improves power of *all* of time series goodness of fit tests referenced above.

Estimators and their properties are described in Section 3.2. Section 3.3 provides specific formulations for linear model goodness of fit and describe our estimators and their properties. Detecting non-linear effects and assessing the fit of GARCH processes is considered in Section 3.4.

Simulation results presented in Section 3.5 show the consistent improved performance of tests based on $\tilde{\rho}_\lambda$. Data applications are in Section 3.6. The chapter is closed with a discussion of the results and future research ideas in Section 3.7.

3.2 Penalized Estimator

Given the data: $x_1, x_2, \ldots, x_n$, we can estimate μ with $\bar{x} = n^{-1}\sum_{t=1}^{n} x_t$. A variety of correlation estimators have been proposed in the literature and most of these can be associated with an objective function which approximates (2.2), with the primary differences between the estimators being how the boundary values (t near 1 and n) are handled. Many time series analysts prefer the sample ACF, since the resulting function is guaranteed to be non-negative definite. Let $y_t = x_t - \bar{x}$, we minimizes $O_h(r|x_1, \ldots, x_n) = \sum_{t=h+1}^{n+h} (y_t - r y_{t-h})^2$ with $y_t = 0$ for $t > n$ to obtain the penalized M-estimator. For a fixed lag m consider estimating the ACF at the first m lags. Let $\boldsymbol{\rho}' = (\rho(1), \ldots, \rho(m))$, $\hat{\boldsymbol{\rho}}' = (\hat{\rho}(1), \ldots, \hat{\rho}(m))$, $\boldsymbol{r}' = (r_1, \ldots, r_m)$, $\boldsymbol{\rho}_\tau' = (\rho_\tau(1), \ldots, \rho_\tau(m))$ be a vector of target values, and $\boldsymbol{\lambda}' = (\lambda_1, \ldots, \lambda_m)$ be a vector of tuning parameters. The penalized ACF estimator is the minimizer of

$$\sum_{h=1}^{m} \left\{ O_h(r_h|x_1, \ldots, x_n) + \lambda_h (\sum_{t=1}^{n} y_t^2)(r_h - \rho_\tau(h))^2 \right\}. \tag{3.1}$$

and the resulting estimator of ACF is

$$\tilde{\rho}_\lambda(h) = w_h \rho_\tau(h) + (1 - w_h)\hat{\rho}(h), \tag{3.2}$$

and the penalized PACF is

$$\tilde{\pi}_\lambda(h) = w_h \pi_\tau(h) + (1 - w_h)\hat{\pi}(h),$$

where $w_h = \lambda_h/(1 + \lambda_h)$, $\hat{\pi}(h)$ is the ordinary sample PACF, and $\pi_\tau(h)$ is a target value. Many correlation estimators proposed in the literature result from special cases of minimizing (3.1) in Table 2.1. Both λ and the target value ρ_τ can be selected from data, in which case $\rho_\tau = \hat{\rho}_\tau$, and $w_h = \hat{w}_h$. For example, taking $\hat{\rho}_\tau = \hat{\rho} + 2bias$ and $\lambda = 1$, results in a bias corrected estimator $\tilde{\rho}_\lambda = \hat{\rho} + bias$, where $bias$ is an additive bias correction term.

Since the sample (partial) autocorrelation estimators are asymptotically unbiased and min-

imize the limiting MSE, it is optimal to select penalized estimators that enjoy the same asymptotic behaviour as that of these estimators. It is simple to ensure that the estimator in (3.2) will retain the same asymptotic distribution as $\hat{\rho}$ (see Theorem 1); either let $\hat{w}_h \to 0$ in probability or as below select $\hat{\rho}_\tau(h)$ to converge sufficiently fast to $\hat{\rho}(h)$. The key is then to select target and tuning parameters from the data to improve the finite sample performance of any statistical procedures based on estimates of the (partial) ACF. Specific choices for these that improve time series goodness of fit are given in Section 3.3.

There is no guarantee that the proposed estimator will provide a non-negative definite solution. For the purpose of time series goodness of fit test, this is not particularly important. A discussion of some techniques for non-negative definite correction is given in Section 2.2.

3.3 Linear Goodness of Fit

Parametric models are frequently used to approximate the correlation structure of a stationary process or of a function of that stationary process. The correlation structure of a time series $\{X_t\}$ is typically modeled using a causal and invertible autoregressive-moving average (ARMA) process of the form

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t \tag{3.3}$$

where p is the autoregressive order, q is the moving average order and $\{\epsilon_t\}$ is an independent and identically distributed (iid) innovation sequence of zero mean random variables with finite variance.

The sample ACF of the residuals $\{\hat{\epsilon}_t\}$:

$$\hat{r}(k) = \frac{\sum_{t=k+1}^{n} \hat{\epsilon}_t \hat{\epsilon}_{t-k}}{\sum_{t=1}^{n} \hat{\epsilon}_t^2} \quad k = 1, 2, \ldots, m,$$

from a fitted ARMA model provide information on the goodness of fit. If the orders p and q are correctly identified, then each of the above correlation coefficients should be approximately equal to zero. However, if the fitted model underestimates the ARMA orders, the values of the autocorrelations should significantly deviate from zero. Most statistical software calculates the Portmanteau statistic

$$Q = n \sum_{k=1}^{m} \frac{n+2}{n-k} \hat{r}^2(k), \tag{3.4}$$

from Ljung and Box (1978). Monti (1994) introduced a test utilizing the residual partial autocorrelations:

$$M = n \sum_{k=1}^{m} \frac{n+2}{n-k} \hat{\pi}^2(k), \tag{3.5}$$

where $\hat{\pi}(k)$ is the residual partial autocorrelation at lag k. Under the null hypothesis, each of these statistics is asymptotically distributed as a chi-squared random variable with $m-(p+q)$ degrees of freedom. Simulations demonstrate that M is more powerful than Q when the fitted model underestimates the order of the moving average component. To increase the power of the ARMA goodness of fit test, Peña and Rodríguez (2002) proposed a statistic based on the determinant of $\hat{\boldsymbol{R}}$, which has a more numerically stable version expressed in terms of the partial correlations (Peña and Rodríguez, 2006). Mahdi and McLeod (2012) generalized the determinant based tests to the multivariate setting and for the univariate case they suggest using

$$\frac{-3n}{2m+1} \log |\widehat{\mathbf{R}}_m|$$

with null asymptotic distribution represented as a linear combination of independent chi-square random variables. Using the relationship between the determinant of $\widehat{\mathbf{R}}_m$ and the partial correlations, this statistic can be calculated as

$$D = n \sum_{k=1}^{m} \frac{-3(m+1-k)}{2m+1} \log(1-\hat{\pi}^2(k)), \tag{3.6}$$

and its null hypothesis asymptotic distribution is approximately chi-square with the degrees of freedom $(3/2)m(m+1)/(2m+1)-(p+q)$. Fisher and Robbins (2018) showed that tests based on the log of the determinant of the correlation matrix increase power in detecting residual correlation. Inspired by high dimensional covariance matrix tests, Fisher and Gallagher (2012) considered a test based on the trace of matrix $\widehat{\mathbf{R}}^2$ and a similar test using partial correlations:

$$Q_W = n \sum_{k=1}^{m} \frac{(n+2)(m-k+1)}{(n-k)m} \hat{r}^2(k) \quad \text{and} \quad M_W = n \sum_{k=1}^{m} \frac{(n+2)(m-k+1)}{(n-k)m} \hat{\pi}^2(k). \tag{3.7}$$

Under the null hypothesis of adequate ARMA fit, the asymptotic distributions of Q_W and M_W can be expressed as a linear combination of chi-square random variables. For computational ease, they

recommend an approximation using gamma distribution with shape and scale parameters

$$\alpha = \frac{3}{4}\frac{m(m+1)^2}{2m^2+3m+1-6m(p+q)} \quad \text{and} \quad \beta = \frac{2}{3}\frac{2m^2+3m+1-6m(p+q)}{m(m+1)}.$$

For a more complete discussion of each of the above test statistics see Fisher and Gallagher (2012).

For each of the above test statistics, the null hypothesis that the fitted ARMA model adequately models the correlation structure is rejected if the test statistic exceeds a critical value; the critical value can be found using an asymptotic distribution or via Monte Carlo methods (Mahdi and McLeod, 2012; Fisher and Gallagher, 2012). Simulations indicate that all of the above tests have bounded Type I error. In terms of power D tends to have higher power than Q or M, while M_W tends to have higher power than D when the order of moving average is underestimated, and Q_W can have higher power than D when the autoregressive order is underestimated.

The power of each of the above test statistics can be improved by replacing $\hat{r}(k)$ with $\tilde{r}_\lambda(k)$ and replacing $\hat{\pi}(k)$ with $\tilde{\pi}_\lambda(k)$, with target values chosen to increase the magnitude of larger values of $\hat{r}^2(k)$ and $\hat{\pi}^2(k)$. The Type I error can be controlled by shrinking smaller magnitude correlations toward zero. Tuning parameters can be chosen via cross validation or by plug in methods. Using Theorem 1 the target values can be chosen to retain the same asymptotic distributions of the test statistics. In the next section specific choices of target values and tuning parameters are given. The simulation results in Section 3.5 show the improved power for each of Q, M, D, Q_W, and M_W, when the sample ACF/PACF are replaced by the penalized M-estimators.

3.3.1 Penalized Estimator for ARMA Fit

Each of the above statistics for testing ARMA fit has representation

$$T_m = n\sum_{k=1}^{m}\omega_k f(\hat{\rho}^2(k)), \tag{3.8}$$

where $\hat{\rho}(k)$ is either $\hat{r}(k)$ or $\hat{\pi}(k)$, f is an increasing function and each summand contributes to the potential rejection of the ARMA fit according to the magnitude of $\hat{\rho}(k)$; for Q,M, Q_W, and M_W, $f(x) = x$, while for D, $f(x) = -\log(1-x^2)$. To improve power, while controlling Type I error, for each test statistic satisfying (3.8) replace $\hat{\rho}(k)$ with penalized M-estimator $\tilde{\rho}_\lambda(k)$. Shrinking apparently insignificant (partial) correlations toward zero will control type I error, while moving

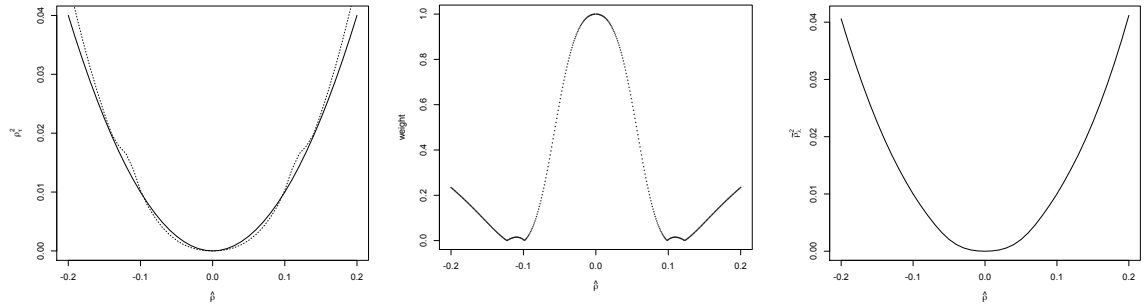
higher magnitude (partial) correlations toward a larger target will increase power. Under the null hypothesis the approximate standard error of the estimated lag k (partial) correlation is $\ell_k = \sqrt{(n-k)/(n(n+2))}$, and correlations less than one standard error from 0 can be shrunk toward zero. Rejection region tests based on the asymptotic distribution of the test statistics reject the adequacy of the fitted ARMA model if $T_m > q_\alpha$, where q_α denotes the $1-\alpha$ quantile of the asymptotic distribution and α is the nominal type I error bound. In order to reject the null hypothesis of model adequacy, at least one summand must exceed $u = q_\alpha/m$, and summands exceeding or near this value support the alternative hypothesis of model under fit. Setting the kth summand in (3.8) equal to u and solving for $\hat{\rho}^2(k)$ results in u_k.

A general form of the target correlation is

$$\rho_\tau(k) = \begin{cases} \hat{\rho}(k)\,(1-a_n) & |\hat{\rho}(k)| < \ell_k \\ \hat{\rho}(k) + Sb_n & \ell_k < |\hat{\rho}(k)| < u_k \\ \hat{\rho}(k)\,(1+c_n) & |\hat{\rho}(k)| > u_k, \end{cases}$$

where $S = \text{sign}(\hat{\rho})$, $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ are positive sequences, with $a_n \to 0$, $\sqrt{n}b_n \to 0$ and $c_n \to 0$ as $n \to \infty$. Each of the Portmanteau test statistics described above, will retain the same null asymptotic distribution when the (partial) correlations are replaced with their penalized versions. Careful choice of these sequences will control type I error while increasing power, and provide stability with respect to the choice of maximum lag m. Suggested principles for selecting target and tuning parameters are:

- To control Type I error:
  - If $\hat{\rho}(k)$ less than one standard error from 0, $|\tilde{\rho}_\lambda(k)| < |\hat{\rho}(k)|$;
  - If $\omega_k f(\hat{\rho}^2(k)) \le q_\alpha/m$, then $\omega_k f(\tilde{\rho}^2_\lambda(k)) \le q_\alpha/m$;
  - If $\omega_k f(\hat{\rho}^2(k)) > q_\alpha/m$, then ρ_τ is larger than $\hat{\rho}(k)$ with the increase being proportional to the distance between $\omega_k f(\hat{\rho}^2(k))$ and q_α/m;
  - $\lambda = 0$, when $\hat{\rho}(k) = \ell_k$ and when $\hat{\rho}(k) = u_k$.
- To increase power:
  - If $\omega_k f(\hat{\rho}(k)^2) > q_\alpha/m$, then ρ_τ is larger than $\hat{\rho}(k)$ with the increase depending on the approximate standard error of $\hat{\rho}(k)$, the sample size n, and the magnitude of $\hat{\rho}(k)$.



(a) Target (b) Weight (c) Penalized Estimator

Figure 3.1: (a) ρ_τ^2, (b) w, and (c) $\tilde{\rho}_\lambda^2$ as functions of $\hat{\rho} = \hat{r}(1)$ for the Ljung-Box test with $\alpha = 0.05$, $n = 100$, and $m = 10$.

Specific choices for target and tuning parameters satisfying these principles are given below. Simulation results in Section 3.5 show the improved performance using these formulae.

For any test satisfying (3.8), consider the target value

$$\rho_\tau(k) = \begin{cases} \hat{\rho}(k)\left(1 + \sqrt{m}(|\hat{\rho}(k)| - \ell_k)\right) & |\hat{\rho}(k)| < \ell_k \\ \hat{\rho}(k) + S_k m \frac{(r-\ell_k)(u_k - r)}{\sqrt{n}(u_k - \ell_k)} & \ell_k < |\hat{\rho}(k)| < u_k \\ \hat{\rho}(k)\left(1 + \frac{m+1-k}{m}(|\hat{\rho}(k)| - u_k)(1 - |\hat{\rho}(k)|)\ell_k\sqrt{nq_\alpha/m}\right) & |\hat{\rho}(k)| > u_k, \end{cases} \tag{3.9}$$

where $\ell_k = \sqrt{(n-k)/(n^2+2n)}$, $u_k = \sqrt{f^{-1}(u/(n\omega_k))}$, $\hat{\rho}(k)$ is the lag k sample (PACF) ACF of the residual sequence, and $S_k = \text{sign}(\hat{\rho}(k))$. The target value depends on the sample size, the number of lags m, the type I error bound α, and the initial estimate of the lag k (partial) correlation. Note that the terms involving m in the target are included to ensure stability with changing m. With the above target, associate tuning parameter:

$$\lambda_k = |(|\hat{\rho}(k)| - \ell_k)(|\hat{\rho}(k)| - u_k)|/(\hat{\rho}^2(k)(1 - \hat{\rho}(k))^2). \tag{3.10}$$

The penalized M-estimator of the ACF and PACF at lag k are calculated as

$$\tilde{r}_\lambda(k) = w_k r_\tau(k) + (1 - w_k)\hat{r}(k) \quad \text{and} \quad \tilde{\pi}_\lambda(k) = w_k \pi_\tau(k) + (1 - w_k)\hat{\pi}(k), \tag{3.11}$$

with $w_k = \lambda_k/(1+\lambda_k)$. For a test statistic of type given by (3.8), let

$$\tilde{T}_m = n\sum_{k=1}^{m} \omega_k f(\tilde{\rho}_\lambda^2(k)), \tag{3.12}$$

where $\tilde{\rho}_\lambda(k)$ is given in (3.11). In particular,

$$\tilde{M} = n\sum_{k=1}^{m} \frac{n+2}{n-k}\tilde{\pi}_\lambda^2(k) \quad \text{and} \quad \tilde{Q} = n\sum_{k=1}^{m} \frac{n+2}{n-k}\tilde{r}_\lambda^2(k),$$

$$\tilde{D} = n\sum_{k=1}^{m} \frac{-3(m+1-k)}{2m+1}\log(1-\tilde{\pi}_\lambda^2(k)), \tag{3.13}$$

$$\tilde{Q}_W = n\sum_{k=1}^{m} \frac{(n+2)(m-k+1)}{(n-k)m}\tilde{r}_\lambda^2(k) \quad \text{and} \quad \tilde{M}_W = n\sum_{k=1}^{m} \frac{(n+2)(m-k+1)}{(n-k)m}\tilde{\pi}_\lambda^2(k).$$

Figure 3.1 demonstrates the target, weight and resulting penalized estimator for $\hat{r}(1)$, when $m = 10$ and the test statisitic $\tilde{Q}$ is to be used with nominal Type I error bound $\alpha = 0.05$. In each panel, $\hat{\rho} = \hat{r}(1)$ is restricted to be in the interval $(-0.2, 0.2)$. Here correlations larger than $\ell_k = 0.099$ are increased in magnitude, and (a) includes a plot of $\hat{r}^2$ vs $\hat{r}$ for reference (solid line).

3.3.2 Asymptotic Distribution

Under the null hypothesis, the plug in choices for $\rho_\tau(k)$ and λ_k given by (3.9) and (3.10), respectively, ensure that $\tilde{T}_m$ and T_m have the same asymptotic distribution. The general asymptotic distribution for $\tilde{T}_m$ is as follows.

Define $\hat{\mathbf{r}}' = (\hat{r}_1, \ldots, \hat{r}_m)$ and $\hat{\boldsymbol{\pi}}' = (\hat{\pi}_1, \ldots, \hat{\pi}_m)$ and using the results of Ljung and Box (1978) and Monti (1994) that $\sqrt{n}\hat{\mathbf{r}}$ and $\sqrt{n}\hat{\boldsymbol{\pi}}$ are asymptotically multivariate normal with mean zero vector and covariance matrix $(\mathbf{I} - \mathbf{Q})$, where $\mathbf{Q} = \mathbf{X}\boldsymbol{\Psi}^{-1}\mathbf{X}'$, $\boldsymbol{\Psi}$ is the information matrix for parameters ϕ and θ, and $\mathbf{X}$ is an $m \times (p+q)$ matrix with elements $\boldsymbol{\phi}'$ and $\boldsymbol{\theta}'$ defined by $1/\phi(B) = \sum \phi_i' B^i$ and $1/\theta(B) = \sum \theta_i' B^i$.

Theorem 2 Let T_m have representation (3.8) and $\tilde{T}_m$ satisfy (3.12). If $f(x^2) = x^2 + o(x^2)$ as $x^2 \to 0$, and $\omega_k \to \tilde{w}_k$ as $n \to \infty$, then under the null hypothesis of an adequately fitted ARMA model

$$T_m \Rightarrow \sum_{k=1}^{m} \gamma_k \chi_k^2 \quad \text{and} \quad \tilde{T}_m \Rightarrow \sum_{k=1}^{m} \gamma_k \chi_k^2,$$

where $\Rightarrow$ denotes convergence in distribution, $\{\chi^2_k\}$ are independent chi-squared random variables with one degree of freedom, and γ_k $(k = 1, \ldots, m)$ are the eigenvalues of $(\mathbf{I} - \mathbf{Q})\mathbf{W}$, where $\mathbf{W}$ is a diagonal matrix with elements $w_{ii} = \tilde{w}_i$ $(i = 1, \ldots, m)$.

Proof: Assume the data is generated by $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t$, for each n the model is fit with the correct orders p and q using $\sqrt{n}$-consistent estimators of the ARMA parameters, and $\hat{\pi}(\cdot)$ and $\hat{r}(\cdot)$ are the sample (partial) autocorrelations of the residuals. Let T_m have representation $n \sum_{k=1}^{m} \omega_k f(\hat{\rho}^2(k))$ and $\tilde{T}_m$ have representation $n \sum_{k=1}^{m} \omega_k f(\tilde{\rho}^2_\lambda(k))$. Denote autocorrelation based tests as $T_m(\hat{\boldsymbol{r}})$ and partial correlation tests with $T_m(\hat{\boldsymbol{\pi}})$. For f and ω_k satisfying the conditions of the theorem, each of the m summands in T_m behaves like

$$n\omega_k f(\hat{r}^2(k)) = n\tilde{w}_k \hat{r}^2(k)(1 + o_p(1)) \quad \text{and} \quad n\omega_k f(\hat{\pi}^2(k)) = n\tilde{w}_k \hat{\pi}^2(k)(1 + o_p(1)),$$

so that

$$T_n(\hat{\boldsymbol{r}}) = n \sum_{k=1}^{m} \tilde{w}_k \hat{r}^2(k) + o_p(1) \quad \text{and} \quad T_n(\hat{\boldsymbol{\pi}}) = n \sum_{k=1}^{m} \tilde{w}_k \hat{\pi}^2(k) + o_p(1).$$

Define

$$\mathbf{W} = \begin{bmatrix} \tilde{w}_1 & 0 & \cdots & 0 \\ 0 & \tilde{w}_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & \cdots & 0 & \tilde{w}_m \end{bmatrix}.$$

as the diagonal matrix asymptotic weights. Then $T_m(\hat{\boldsymbol{r}})$ and $T_m(\hat{\boldsymbol{\pi}})$ are asymptotically expressed as quadratic forms

$$T_m(\hat{\boldsymbol{r}}) \simeq n\hat{\mathbf{r}}'\mathbf{W}\hat{\mathbf{r}} \quad \text{and} \quad T_m(\hat{\boldsymbol{\pi}}) \simeq n\hat{\boldsymbol{\pi}} W \hat{\boldsymbol{\pi}} \quad \text{as } n \to \infty,$$

respectively, where A' denotes the transpose operation on vector/matrix A and $\hat{\mathbf{r}}$ $(\hat{\boldsymbol{\pi}})$ is the $m \times 1$ vector of the autocorrelations (partial autocorrelations) from lag 1 to m. From the results in Box (1954) both quadratic forms will be distributed as

$$\sum_{k=1}^{m} \lambda_k \chi^2_k \tag{3.14}$$

where each χ^2_k are independently distributed chi-squared random variables with one degree of free-

dom, and the λ_k are the m real nonzero characteristic roots of the matrix $(\mathbf{I}-\mathbf{Q})\mathbf{W}$ where $\mathbf{I}-\mathbf{Q}$ is the $m \times m$ covariance matrix of both $\sqrt{n}\hat{\mathbf{r}}$ and $\sqrt{n}\hat{\pi}$. $\mathbf{Q}$ is the $m \times m$ information matrix for ϕ and θ. Box and Pierce (1970) and McLeod (1978) approximate the matrix $\mathbf{Q}$ by the projection matrix $\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ when m is moderately high.

Let $\tilde{r}_\lambda(k)$ and $\tilde{\pi}_\lambda(k)$ be calculated using

$$\tilde{r}_\lambda(k) = w_k r_\tau(k) + (1-w_k)\hat{r}(k) \quad \text{and} \quad \tilde{\pi}_\lambda(k) = w_k \pi_\tau(k) + (1-w_k)\hat{\pi}(k), \tag{3.15}$$

with the target and tuning parameters given as

$$\rho_\tau(k) = \begin{cases} \hat{\rho}(k)\left(1+\sqrt{m}(|\hat{\rho}(k)|-\ell_k)\right) & |\hat{\rho}(k)| < \ell_k \\ \hat{\rho}(k) + S_k m \frac{(r-\ell_k)(u_k-r)}{\sqrt{n}(u_k-\ell_k)} & \ell_k < |\hat{\rho}(k)| < u_k \\ \hat{\rho}(k)\left(1+\frac{m+1-k}{m}(|\hat{\rho}(k)|-u_k)(1-|\hat{\rho}(k)|)\ell_k\sqrt{nq_\alpha/m}\right) & |\hat{\rho}(k)| > u_k, \end{cases}$$

where $\ell_k = \sqrt{(n-k)/(n^2+2n)}$, $u_k = \sqrt{f^{-1}(u/(n\omega_k))}$, $\hat{\rho}(k)$ is the lag k sample (PACF) ACF of the residual sequence, and $S_k = \text{sign}(\hat{\rho}(k))$, and the tuning parameter

$$\lambda_k = |(|\hat{\rho}(k)|-\ell_k)(|\hat{\rho}(k)|-u_k)|/(\hat{\rho}^2(k)(1-\hat{\rho}(k))^2).$$

respectively. Then for each k

$$\sqrt{n}(\hat{\rho}(k)-\rho_\tau(k)) = \sqrt{n}\begin{cases} \hat{\rho}(k)\left(\sqrt{m}(|\hat{\rho}(k)-\ell_k)\right) & |\hat{\rho}(k)| < \ell \\ Sm\frac{(|\hat{\rho}(k)|-\ell_k)(u_k-|\hat{\rho}(k)|)}{\sqrt{n}(u_k-\ell_k)} & \ell_k < |\hat{\rho}(k)| < u_k \\ \hat{\rho}(k)\left(\log(n)(|\hat{\rho}(k)|-u)(1-|\hat{\rho}(k)|)/\sqrt{n}\right) & |\hat{\rho}(k)| > u_k, \end{cases}$$

where as $n \to \infty$: $\sqrt{n}\hat{\rho}$ is asymptotically normally distributed, $\hat{\rho} \to 0$ in probability, $\ell_k \to 0$, and $u_k \to 0$. Each term on the right hand side of the above equation converges to 0 in probability. By Theorem 1, $\sqrt{n}\tilde{\mathbf{r}}_\lambda$ and $\sqrt{n}\hat{\mathbf{r}}$ have the same limiting distribution, and $\sqrt{n}\tilde{\boldsymbol{\pi}}_\lambda$ and $\sqrt{n}\hat{\boldsymbol{\pi}}$ have the same limiting distribution. Note that $T_m(\tilde{\boldsymbol{r}}_\lambda)$ and $T_m(\tilde{\boldsymbol{\pi}}_\lambda)$ are asymptotically expressed as quadratic forms

$$T_m(\hat{\boldsymbol{r}}) \simeq n\tilde{\mathbf{r}}_\lambda'\mathbf{W}\hat{\mathbf{r}}_\lambda \quad \text{and} \quad T_m(\hat{\boldsymbol{\pi}}) \simeq n\tilde{\boldsymbol{\pi}}_\lambda\mathbf{W}\tilde{\boldsymbol{\pi}}_\lambda \quad \text{as } n \to \infty.$$

If follows that T_m and $\tilde{T}_m$ have the same asymptotic distribution.

Corollary 1 Let $\{X_t\}$ be generated by (3.3). Under the null hypothesis of an adequately fitted model,

1. Q and $\tilde{Q}$ have the same asymptotic distribution.
2. Q_W and $\tilde{Q}_W$ have the same asymptotic distribution.
3. M and $\tilde{M}$ have the same asymptotic distribution.
4. M_W and $\tilde{M}_W$ have the same asymptotic distribution.
5. D and $\tilde{D}$ have the same asymptotic distribution.

The improved power and controlled Type I error of the proposed test statistics for a variety of models will be given in Section 3.5. Here a toy example demonstrates the improved local power as a function of n.

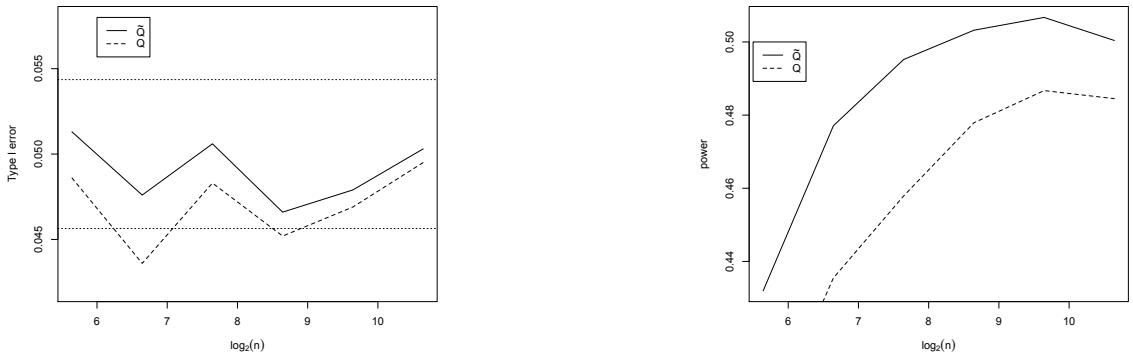
Example 1 *Let $\{X_t\}_{t=1}^n$ be generated by difference equation*

$$X_t = \phi_n X_{t-2} + \epsilon_t,$$

where $\phi_n = 3/\sqrt{n}$ and $\{\epsilon_t\}$ is an IID standard normal sequence. For each $n = 50$, 100, 200, 400, 800, and 1600, data was simulated from the above model, the model was correctly fit via Gaussian maximum likelihood using a second order autoregressive model, and the model was under fit with a first order autoregressive model. This process was repeated $10,000$ times. Simulated type I error as a function of n ($log_2(n)$ scale) based on Q and $\tilde{Q}$ is given in the first panel of Figure 3.2; the horizontal lines correspond to $.05 \pm 2sd$, where sd=$\sqrt{0.05(0.95)/10000}$ is the standard error from the binomial distribution. The second panel of Figure 3.2 shows simulated power for Q and $\tilde{Q}$ respectively.

Calculating p-values. The proposed test statistics depend on a nominal bound on Type I error, α. For each fixed α, we reject the model fit if $\tilde{T}_m(\alpha) > q_\alpha$, with (partial) autocorrelation estimated using (3.11), and target and tuning parameters calculated using q_α. We consider the p-value as the largest α for which we would reject the model fit:

$$pvalue = sup\{\alpha : \tilde{T}_m(\alpha) > q_\alpha\}. \tag{3.16}$$



(a) Type I error

(b) Power

Figure 3.2: Type I error and power as function of n in Example 1; Test statistics calculated using $m = 10$. (a) Simulated type I error for Q and $\tilde{Q}$ plotted vs $log_2(n)$. (b) Simulated power for Q and $\tilde{Q}$ plotted vs $log_2(n)$.

3.4 Non-linear Models

Linear models cannot capture all observed data features. In financial time series, empirical evidence motivates modeling the correlation structure of a non-linear function of the process (Engle, 1982). Many authors have considered nonlinear models of the type:

$$\epsilon_t = g(h_t)\eta_t \tag{3.17}$$

where η_t are iid with mean zero and variance 1 and $\{h_t\}$ follows an ARMA type recursion. For example, the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process of Bollerslev (1986) takes $g(x) = \sqrt{x}$, while the stochastic volatility model (SV) of Taylor (1986) assumes $g(x) = \exp(x)$. For these models a nonlinear function of $\{\epsilon_t\}$ has the correlation structure of an ARMA process. Here we consider tests to detect the presence of nonlinear effects, as well as goodness-of-fit of a nonlinear process.

3.4.1 Detecting nonlinear effects

Transformed residuals of a fitted model can be used to detect nonlinear effects (Pérez and Ruiz, 2003). Consider the autocorrelation of the transformed residuals:

$$\hat{r}^*(k) = \frac{\sum_{t=k+1}^{n}(g(\hat{\epsilon}_t) - \bar{g}(\hat{\epsilon}_t))(g(\hat{\epsilon}_{t-k}) - \bar{g}(\hat{\epsilon}_t))}{\sum_{t=1}^{n}(g(\hat{\epsilon}_t) - \bar{g}(\hat{\epsilon}_t))^2}, \tag{3.18}$$

where $\bar{g}(\hat{\epsilon}_t) = \sum g(\hat{\epsilon}_t)/n$, and let $\hat{\pi}^*(k)$ denote the lag k sample partial correlation of the transformed residuals. The vector of autocorrelation coefficients based on the squared residuals of a fitted ARMA model with iid innovations, are asymptotically normally distributed with mean zero and unit covariance matrix (McLeod and Li, 1983b). The (partial) autocorrelation coefficients of absolute residuals or the log of the squared residuals can be used detect GARCH or SV structure (Peña and Rodríguez, 2006; Fisher and Gallagher, 2012).

Each of the statistics for ARMA goodness of fit from the previous section can be used to test for non-linear effects (McLeod and Li, 1983b; Peña and Rodríguez, 2002; Fisher and Gallagher, 2012). McLeod and Li (1983b) suggest a Ljung Box type test statistic,

$$Q^* = n(n+2)\sum_{k=1}^{m}\frac{\hat{r}^{*2}(k)}{n-k}, \tag{3.19}$$

where the $*$ represents the squared residuals, and show that Q^* is asymptotically distributed as chi-squared with m degrees of freedom. A version of Monti's statistic to detect non-linear effects would be

$$Q^* = n(n+2)\sum_{k=1}^{m}\frac{\hat{\pi}^{*2}(k)}{n-k}, \tag{3.20}$$

where $\hat{\pi^*}$ denotes the partial correlation of the transformed residuals. For a fixed function $g(\cdot)$, consider a general class of test statistics with form:

$$T_m^* = n\sum_{k=1}^{m}\omega_k f(\hat{\rho}^{*2}(k)), \tag{3.21}$$

where $\hat{\rho}^*(k)$ is either $\hat{r}^*(k)$ or $\hat{\pi}^*(k)$, f is an increasing function and each summand contributes to the potential rejection of the ARMA fit according to the magnitude of $\hat{\rho}^*(k)$; for Q^*,M^*, Q_W^*, and M_W^*, $f(x) = x$, while for D^*, $f(x) = -\log(1-x^2)$.

For any test statistic satisfying (3.21), there is an associated test using the penalized corre-

lation estimator:

$$\tilde{T}_m^* = n \sum_{k=1}^{m} \omega_k f(\tilde{\rho}_\lambda^{*2}(k)), \tag{3.22}$$

where $\tilde{\rho}_\lambda^*(k)$ is calculated as in Section 3.3, but with $\hat{\rho}(k)$ replaced with $\hat{\rho}^*(k)$, and using quantiles from the distribution of T_m^*.

Theorem 3 Let $\{\hat{\epsilon}_t\}$ be residuals from (correctly) fitted ARMA model (3.3). If T_m^* in (3.21) and $\tilde{T}_m^*$ in (3.22) are computed from the squared, absolute or log of the squared residuals, and $f(x^2) = x^2 + o(x^2)$ as $x^2 \to 0$, and $\omega_k \to \gamma_k$ as $n \to \infty$, then

$$T_m^* \Rightarrow \sum_{k=1}^{m} \gamma_k \chi_k^2 \quad \text{and} \quad \tilde{T}_m^* \Rightarrow \sum_{k=1}^{m} \gamma_k \chi_k^2,$$

where $\Rightarrow$ denotes convergence in distribution and $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom.

Proof: We give the proof for squared residuals. The proofs for absolute residuals, and of the log of squared-residuals are similar. From the result in McLeod and Li (1983) the autocorrelations of the squared residuals are asymptotically normally distribution as $\sqrt{n}\hat{\mathbf{r}}(\hat{\epsilon}_t^2) \to N(0, \mathbf{I}_m)$ where $\mathbf{I}_m$ is an $m \times m$ identity matrix. Applying this result to the results in Monti (1994) provides the asymptotic distribution $\sqrt{n}\hat{\boldsymbol{\pi}}(\hat{\epsilon}_t^2) \to N(0, \mathbf{I}_m)$. $T_m^*(\hat{\boldsymbol{r}})$ and $T_m^*(\hat{\boldsymbol{\pi}})$ can be expressed as quadratic forms as in the proof of Theorem 2. Theorem 1 establishes that T_m^* and $\tilde{T}_m^*$ have the same asymptotic distribution.

3.4.2 GARCH and ARCH Goodness-of-fit

In time series econometrics, the ARCH models of Engle (1982) and GARCH models of Bollerslev (1986) have received much attention. The GARCH(b, a) model is given by $g(x) = \sqrt{x}$ in (3.17) where $\{\eta_t\}$ is iid with zero mean and variance of unity and

$$h_t = \omega + \sum_{i=1}^{b} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{a} \beta_j h_{t-j}.$$

Ideally, after fitting the parameters for the GARCH process, checking for the adequacy of that model should follow. Higgins and Bera (1992) suggested a test based on the sum of squared residual sample autocorrelations, but Li and Mak (1994) showed that the statistic does not converge to a chi-squared distribution asymptotically when constructed with the squared residuals. Instead

they proposed the statistic based on the autocorrelation function developed by the *standardized* sample squared residuals,

$$\breve{r}(k) = \frac{\sum_{t=k+1}^{n}(\hat{\epsilon}_t^2/\hat{h}_t - \bar{\epsilon})(\hat{\epsilon}_{t-k}^2/\hat{h}_{t-k} - \bar{\epsilon})}{\sum_{t=1}^{n}(\hat{\epsilon}_t^2/\hat{h}_t - \bar{\epsilon})^2}$$

where $\bar{\epsilon} = (1/n)\sum \hat{\epsilon}_t^2/\hat{h}_t$ and $\hat{h}_t$ are the sample conditional variances. The statistic

$$L(m) = n\breve{\mathbf{r}}'\hat{\mathbf{V}}^{-1}\breve{\mathbf{r}} \tag{3.23}$$

will be asymptotically distributed as a chi-squared random variable with m degrees of freedom. The statistic requires $\hat{\mathbf{V}}$ to be a consistent estimator for the covariance matrix of $\breve{\mathbf{r}}$.

Li and Mak (1994) also show that for an ARCH(b) model (GARCH($b, 0$)) the residual autocorrelations $\breve{r}(k)$ for $k = b+1, \ldots, m$ are asymptotically iid standard normal, and propose the modified statistic

$$L(b, m) = n\sum_{k=b+1}^{m}\breve{r}^2(k). \tag{3.24}$$

Their simulation study shows only modest improvement using $L(m)$ compared to $L(b, m)$ when the data follows an ARCH process. They suggest a practitioner may prefer $L(b, m)$ for its simplicity in checking the adequacy of a fitted ARCH model.

Tse (2002) also proposed a statistic using lagged squared standardized residuals, which can be calculated via a recursive method. Tsui (2004) gave simulation results showing that the statistics of Li and Mak (1994) and Tse (2002) are comparable and more powerful than the general regression based test from Wooldridge (1991). Li and Li (2005) derived the distribution of $\breve{\mathbf{r}}$ using an absolute deviations approach for fitting the GARCH process and suggest a statistic akin to $L(m)$ in (3.23). Fisher and Gallagher (2012) showed improved power in detecting ARCH under fit, using a weighted version of $L_{b,m}$:

$$L_W(b, m) = n\sum_{k=b+1}^{m}\frac{m-k+(b+1)}{m}\breve{r}^2(k). \tag{3.25}$$

the distribution can be approximated with a gamma distribution with shape and scale parameters:

$$\alpha = \frac{3}{4}\frac{(m-b)(m+b+1)^2}{2m^2+3m+2mb+2b^2+3b+1} \quad \text{and} \quad \beta = \frac{2}{3}\frac{2m^2+3m+2mb+2b^2+3b+1}{m(m+b+1)}.$$

Each of the above statistics can be improved using the penalized M-estimator of autocorrelation, but we focus on detecting ARCH under fit.

Consider statistics of the form

$$L_{\mathbf{a}}(b,m) = n \sum_{k=b+1}^{m} \omega_k f(\breve{r}^2(k)), \tag{3.26}$$

where f is an increasing function and $\mathbf{a}$ is a vector of weights. The penalized correlation version is

$$\tilde{L}_{\mathbf{a}}(b,m) = n \sum_{k=b+1}^{m} \omega_k f(\breve{r}_\lambda^2(k)), \tag{3.27}$$

where $\breve{r}_\lambda(k)$ is found using (3.11) with q_α coming from the asymptotic distribution of $L_{\mathbf{a}}(b,m)$.

Theorem 4 Let $L_{\mathbf{a}}(b,m)$ have representation (3.26) and $\tilde{L}_{\mathbf{a}}(b,m)$ satisfy (3.27). If $f(x^2) = x^2 + o(x^2)$ as $x^2 \to 0$, and $\omega_k \to \gamma_k$ as $n \to \infty$, then under the null hypothesis of an adequately fitted ARCH model

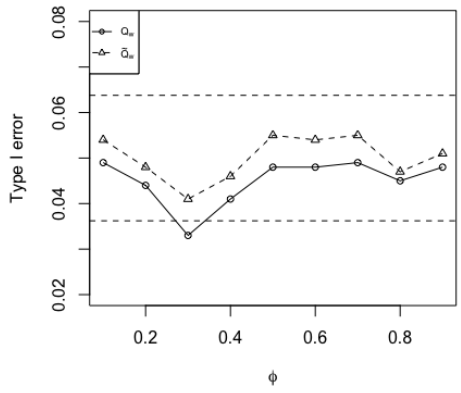
$$L_{\mathbf{a}}(b,m) \Rightarrow \sum_{k=b+1}^{m} \gamma_k \chi_k^2 \quad \text{and} \quad \tilde{L}_{\mathbf{a}}(b,m) \Rightarrow \sum_{k=b+1}^{m} \gamma_k \chi_k^2,$$

where $\Rightarrow$ denotes convergence in distribution, $\{\chi_k^2\}$ are independent chi-squared random variables with one degree of freedom.

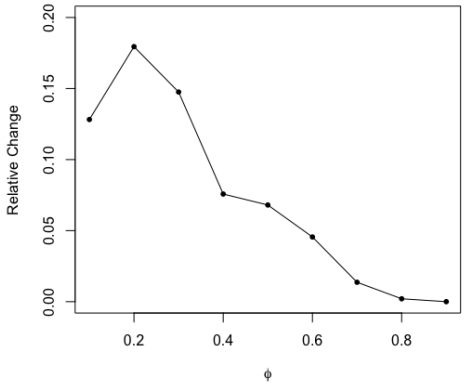
Proof: From the result in Li and Mak (1994), for an ARCH(b) model, the autocorrelations of the standardized squared residuals are asymptotically normally distributed as $N(0, \mathbf{I}_{m-(b+1)})$ where $\mathbf{I}_{m-(b+1)}$ is an $(m-(b+1)) \times (m-(b+1))$ identity matrix. Expressing $L_{\boldsymbol{\omega}}(b,m)$ as a quadratic form and applying Theorem 1 establishes the result.

3.5 Simulation Study

Performance of the proposed tests is compared to tests using the sample ACF/PACF. Following methods from Fisher and Gallagher (2012), where appropriate approximate quantiles from Gamma distributions are used for their test statistics. Simulation were implemented in the statistical language R; source code is provided in the functions Penalized.cor.gof, Penalized.Box.test, Penalized.MahdiMcLeod.test and Penalized.LM.test using the `PenalizedPortTest` package in R. The descriptions of functions in our R-package are given in Appendix B. Simulation results con-



(a) Type I error, AR(1) fitted



(b) Relative Change in Power, MA(1) fitted

Figure 3.3: Empirical size (a) and Power change (b) for $\tilde{Q}_W$, Q_W, and D tests when AR(1) or MA(1) fitted to a series of length $n = 100$, $m = 20$ and parameter $\phi = 0.1, 0.2, ...0.9$ generated by AR(1) process.

sistently tell the same story as seen in Figure 3.2: the penalized based tests control type I error, while improving power over those based on the ordinary sample ACF/PACF. Results are based on nominal type I error level $\alpha = 0.05$, and tables containing power results have the highest empirical power boldfaced in each simulation setting. In every case: (a) the most powerful test uses the penalized ACF or PACF, and (b) tests based on the penalized ACF/PACF have higher power than their sample ACF/PACF counterparts.

3.5.1 Size and Power Studies on Fitted ARMA Models

Results are first presented for linear models. Figure 3.3 illustrates the comparison of type I error and power between $\tilde{Q}_W$ and Q_W for several values of AR(1) parameters ϕ, using 1000 replications of sample size $n = 100$ and $m = 20$. Figure 3.3a shows that both tests have reasonable empirical size, while Figure 3.3b shows the relative change of the power levels $((Power_{\tilde{Q}_W} - Power_{Q_W})/Power_{Q_W})$ for varied parameter ϕ in AR(1) model when $MA(1)$ is fitted. The penalized test improves over the non-penalized version, and provides about a 10% improvement in power when an AR(1) with lag one correlation between 0.1 and 0.3 is incorrectly fit with an MA(1) model.

Table 3.1 provides the empirical significance level of the statistics under AR(1) and MA(1) models. In each case, 1000 replications of sample size $n = 100$ where generated using ϕ or $\theta =$

$0.1, 0.3, 0.5, 0.7, 0.9$. The new tests have slightly higher simulated size for $m = 10, 20$ than their non-penalized counterparts, but Type I errors are controlled. As observed elsewhere, the Ljung and Box (1978) statistic can have slightly elevated empirical size, while the chi-square approximation for the Mahdi and McLeod (2012) statistic provides conservative Type I errors (Mahdi and McLeod, 2012; Fisher and Gallagher, 2012).

Table 3.1: Empirical size at 5% for AR(1) and MA(1) models, $n = 100$, $m = 10, 20$

m	ϕ/θ	$\tilde{Q}$	Q	$\tilde{M}$	M	$\tilde{Q}_W$	Q_W	$\tilde{M}_W$	M_W	$\tilde{D}$	D
						Fitted AR(1)					
10	0.1	0.058	0.055	0.059	0.055	0.041	0.036	0.041	0.039	0.042	0.039
	0.3	0.050	0.049	0.058	0.054	0.033	0.03	0.034	0.033	0.034	0.027
	0.5	0.058	0.056	0.049	0.047	0.040	0.039	0.029	0.025	0.028	0.024
	0.7	0.043	0.038	0.051	0.049	0.031	0.030	0.036	0.032	0.036	0.029
	0.9	0.064	0.061	0.065	0.063	0.062	0.057	0.059	0.053	0.057	0.050
20	0.1	0.065	0.065	0.071	0.066	0.054	0.049	0.047	0.040	0.039	0.032
	0.3	0.061	0.059	0.048	0.044	0.041	0.033	0.036	0.031	0.024	0.021
	0.5	0.068	0.064	0.052	0.052	0.055	0.048	0.049	0.045	0.037	0.026
	0.7	0.065	0.062	0.047	0.043	0.055	0.049	0.040	0.032	0.029	0.022
	0.9	0.054	0.052	0.035	0.036	0.051	0.048	0.039	0.033	0.034	0.027
						Fitted MA(1)					
10	0.1	0.058	0.055	0.068	0.066	0.037	0.036	0.047	0.043	0.045	0.040
	0.3	0.051	0.051	0.053	0.047	0.037	0.031	0.041	0.036	0.04	0.036
	0.5	0.048	0.047	0.058	0.050	0.036	0.032	0.038	0.034	0.035	0.032
	0.7	0.064	0.058	0.061	0.058	0.046	0.041	0.049	0.044	0.047	0.041
	0.9	0.070	0.067	0.060	0.057	0.064	0.057	0.061	0.053	0.058	0.046
20	0.1	0.059	0.058	0.050	0.047	0.046	0.043	0.047	0.037	0.031	0.027
	0.3	0.067	0.064	0.048	0.047	0.050	0.048	0.042	0.035	0.030	0.027
	0.5	0.059	0.057	0.049	0.048	0.042	0.042	0.034	0.028	0.024	0.015
	0.7	0.052	0.046	0.050	0.041	0.039	0.035	0.033	0.027	0.022	0.019
	0.9	0.083	0.080	0.060	0.057	0.065	0.060	0.059	0.053	0.044	0.034

Power of the proposed penalized statistics are compared with the statistic Q from Ljung and Box (1978), M from Monti (1994), Q_W and M_W from Fisher and Gallagher (2012), and D

from Mahdi and McLeod (2012), via a simulation study similar to computational studies appearing previously in the literature (Monti, 1994; Peña and Rodríguez, 2002, 2006; Fisher and Gallagher, 2012).

Table 3.2: Power levels of the tests for ARMA(2,2) models, fitted by AR(1) model, $n = 100$, $m = 10, 20$

$(\phi_1, \phi_2, \theta_1, \theta_2)$	$\tilde{Q}$	Q	$\tilde{M}$	M	$\tilde{Q}_W$	Q_W	$\tilde{M}_W$	M_W	$\tilde{D}$	D
					m=10					
(-, -, 0.5, -)	0.294	0.282	0.338	0.314	0.369	0.343	**0.421**	0.402	0.420	0.394
(-, -, 0.8, -)	0.798	0.747	0.966	0.959	0.914	0.901	**0.990**	**0.990**	0.989	0.988
(-, -, 0.6, -0.3)	0.814	0.761	0.994	0.992	0.930	0.909	**0.998**	0.997	0.997	0.997
(0.1, 0.3, -, -)	0.456	0.420	0.455	0.404	0.543	0.517	0.558	0.533	**0.564**	0.532
(1.1, -0.35, - , -)	0.731	0.711	0.727	0.706	0.821	0.806	**0.824**	0.812	**0.824**	0.807
(0.7, -, 0.4, -)	0.554	0.527	0.627	0.603	0.680	0.664	0.765	0.746	**0.766**	0.746
(0.7, -, 0.9, -)	0.992	0.989	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**
(0.4, -, 0.6, -0.3)	0.883	0.847	0.997	0.997	0.965	0.958	**0.999**	0.998	0.998	0.998
(0.7, -, -0.7, 0.15)	0.186	0.178	0.180	0.163	**0.202**	0.188	0.196	0.178	0.195	0.175
(0.7, 0.2, -0.5, -)	0.779	0.768	0.772	0.756	0.829	0.816	**0.835**	0.829	0.832	0.823
(0.7, 0.2, 0.5, -)	0.392	0.356	0.487	0.453	0.554	0.507	0.644	0.628	**0.645**	0.623
(0.9, -0.4, -1.2, 0.3)	0.764	0.716	0.978	0.974	0.876	0.852	**0.987**	0.985	**0.987**	0.985
					m=20					
(-, -, 0.5, -)	0.236	0.223	0.227	0.198	0.297	0.269	**0.320**	0.291	0.298	0.263
(-, -, 0.8, -)	0.635	0.579	0.883	0.848	0.824	0.786	**0.977**	0.970	0.971	0.965
(-, -, 0.6, -0.3)	0.704	0.644	0.953	0.936	0.853	0.813	**0.991**	0.988	0.988	0.984
(0.1, 0.3, -, -)	0.385	0.353	0.327	0.264	**0.499**	0.463	0.472	0.423	0.463	0.387
(1.1, -0.35, -, -)	0.688	0.656	0.616	0.585	**0.805**	0.788	0.772	0.756	0.752	0.725
(0.7, -, 0.4, -)	0.467	0.440	0.496	0.455	0.621	0.593	**0.678**	0.647	0.659	0.623
(0.7, -, 0.9, -)	0.948	0.924	0.999	0.999	0.994	0.991	**1.000**	**1.000**	**1.000**	**1.000**
(0.4, -, 0.6, -0.3)	0.744	0.688	0.991	0.985	0.914	0.883	**0.998**	**0.998**	**0.998**	0.997
(0.7, -, -0.7, 0.15)	0.169	0.161	0.123	0.114	0.18	**0.168**	0.159	0.140	0.142	0.131
(0.7, 0.2, -0.5, -)	0.656	0.642	0.630	0.596	0.769	0.754	**0.770**	0.755	0.753	0.731
(0.7, 0.2, 0.5, -)	0.330	0.298	0.328	0.298	0.428	0.401	**0.500**	0.46	0.469	0.429
(0.9, -0.4, -1.2, 0.3)	0.605	0.558	0.914	0.900	0.772	0.735	**0.984**	0.976	0.972	0.957

Table 3.2 and Table 3.3 show the power of the ten statistics for 24 different ARMA(2,2) models when $AR(1)$ or $MA(1)$ models are fitted. In each case, 1000 time series of sample size $n = 100$ were generated and the power was computed at $m = 10$ and $m = 20$. The penalized

tests $\tilde{Q}$, $\tilde{M}$, $\tilde{Q}_W$, $\tilde{M}_W$ and $\tilde{D}$ proposed in this chapter almost invariably have more power than the corresponding tests using the sample ACF/PACF, and never have less power. Furthermore, in each simulation setting, one of the proposed statistics, $\tilde{Q}_W$, $\tilde{M}_W$, or $\tilde{D}$, is always the most powerful amongst all considered tests. It appears that penalizing improves performance for *all* commonly used ARMA goodness of fit tests.

Table 3.3: Power levels of the tests for ARMA(2,2) models, fitted by MA(1) model, $n = 100$, $m = 10, 20$

$(\phi_1, \phi_2, \theta_1, \theta_2)$	$\tilde{Q}$	Q	$\tilde{M}$	M	$\tilde{Q}_W$	Q_W	$\tilde{M}_W$	M_W	$\tilde{D}$	D
					m=10					
(0.5, -, -, -)	0.291	0.265	0.254	0.240	0.357	**0.336**	0.346	0.322	0.355	0.319
(0.8, -, - ,-)	0.982	0.978	0.976	0.972	0.989	0.989	**0.990**	0.989	**0.990**	0.989
(1.1, -0.35, -, -)	0.998	0.997	0.995	0.995	**1.000**	**1.000**	**1.000**	0.999	**1.000**	0.999
(-, -, -0.8, 0.5)	0.858	0.833	0.949	0.944	0.947	0.939	**0.980**	0.978	0.978	0.977
(-, -, 0.6, -0.3)	0.396	0.364	0.463	0.430	0.540	0.520	**0.618**	0.600	0.615	0.592
(0.5, -, 0.7, -)	0.893	0.882	0.868	0.848	**0.944**	0.943	0.939	0.933	0.942	0.934
(-0.5, -, -0.7, -)	0.897	0.883	0.898	0.880	**0.956**	0.951	0.954	0.944	0.955	0.943
(0.3, -, -0.8, -0.5)	0.639	0.605	0.762	0.744	0.765	0.745	**0.841**	0.833	**0.841**	0.823
(0.8, -, 0.5, -0.3)	0.987	0.982	0.973	0.969	**0.995**	0.995	0.993	0.990	0.993	0.990
(1.2, -0.5, -0.9, -)	0.487	0.464	**0.727**	0.709	0.472	0.458	0.657	0.642	0.638	0.615
(0.3, -0.2, 0.7, -)	0.264	0.251	0.275	0.263	0.311	0.294	**0.348**	0.327	0.338	0.319
(0.9, -0.4, -1.2, 0.3)	0.813	0.778	0.942	0.937	0.912	0.898	**0.966**	**0.966**	0.965	0.961
					m=20					
(0.5, -, -, -)	0.285	0.269	0.210	0.187	0.321	0.304	0.309	0.277	0.287	0.248
(0.8, -, -, -)	0.967	0.963	0.950	0.933	**0.987**	0.983	0.981	0.976	0.977	0.972
(1.1, -0.35, -, -)	0.989	0.984	0.985	0.983	**0.998**	0.996	0.997	0.996	0.996	0.996
(-, -, -0.8, 0.5)	0.748	0.717	0.869	0.846	0.905	0.890	**0.966**	0.962	0.961	0.948
(-, -, 0.6, -0.3)	0.377	0.355	0.393	0.361	0.475	0.443	**0.559**	0.530	0.526	0.493
(0.5, -, 0.7, -)	0.820	0.794	0.767	0.730	**0.911**	0.900	0.884	0.870	0.879	0.859
(-0.5, -, -0.7, -)	0.853	0.827	0.799	0.756	0.929	0.918	0.911	0.901	0.903	0.888
(0.3, -, -0.8, -0.5)	0.510	0.480	0.606	0.571	0.669	0.632	**0.784**	0.762	0.756	0.732
(0.8, -, 0.5, -0.3)	0.974	0.968	0.946	0.929	**0.992**	0.991	0.976	0.973	0.976	0.970
(1.2, -0.5, -0.9, -)	0.406	0.383	0.598	0.581	0.479	0.449	**0.691**	0.661	0.643	0.609
(0.3, -0.2, 0.7, -)	0.237	0.231	0.216	0.202	**0.286**	0.262	0.279	0.261	0.242	0.223
(0.9, -0.4, -1.2, 0.3)	0.648	0.610	0.847	0.826	0.810	0.780	**0.935**	0.929	0.924	0.916

3.5.2 Detecting Nonlinear Processes

To detect nonlinearity in time series data using the proposed statistics, several nonlinear models are considered and the performance is compared to statistics using the sample (partial) autocorrelation estimator. Typically tests to check the nonlinearity assumption are based on the squared residuals, the absolute residuals, or the log of the squared residuals. Pérez and Ruiz (2003) find that statistics based on the squared and absolute residuals are more powerful than these using log of the squared residuals. Results in Fisher and Gallagher (2012) show that the power of statistics using squared residuals tends to be larger than that using absolute residuals in most cases. Here the size and power of the statistics $\tilde{Q}^*$, Q^*, $\tilde{M}^*$, M^*, $\tilde{Q}^*_W$, Q^*_W, $\tilde{M}^*_W$, M^*_W, $\tilde{D}^*$, and D^* using squared residuals are studied. The simulation settings of Fisher and Gallagher (2012) are followed.

Table 3.4 provides empirical sizes for the ten statistics of interest when a sample of $n = 204$ is taken from various AR(1) processes and properly fit. The nominal level 5% sizes are reported for statistics at lag $m = 24$. As seen in the table, the size of the proposed statistics are similar to those of the corresponding tests without penalizing.

Table 3.4: Empirical sizes at size 5% for AR(1) models, n=204, m=24

ϕ	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}^*_W$	Q^*_W	$\tilde{M}^*_W$	M^*_W	$\tilde{D}^*$	D^*
0.1	0.068	0.067	0.058	0.050	0.062	0.056	0.058	0.053	0.049	0.041
0.3	0.051	0.049	0.037	0.034	0.044	0.041	0.042	0.040	0.036	0.029
0.5	0.052	0.049	0.039	0.037	0.043	0.04	0.043	0.039	0.032	0.030
0.7	0.052	0.046	0.043	0.042	0.046	0.045	0.042	0.039	0.034	0.032
0.9	0.066	0.064	0.053	0.050	0.073	0.069	0.060	0.054	0.047	0.044

Consider eight nonlinear models to compare the powers of the testing statistics,

NL-1: $Y_t = \epsilon_t - 0.4\epsilon_{t-1} + 0.3\epsilon_{t-2} + 0.5\epsilon_t\epsilon_{t-2}$

NL-2: $Y_t = \epsilon_t - 0.3\epsilon_{t-1} + 0.2\epsilon_{t-2} + 0.4\epsilon_t\epsilon_{t-2} - 0.25\epsilon^2_{t-2}$

NL-3: $Y_t = 0.4Y_{t-1} - 0.3Y_{t-2} + 0.5Y_{t-1}\epsilon_{t-1} + \epsilon_t$

NL-4: $Y_t = 0.4Y_{t-1} - 0.3Y_{t-2} + 0.5Y_{t-1}\epsilon_{t-1} + 0.8\epsilon_{t-1} + \epsilon_t$

NL-5: $Y_t = 0.4Y_{t-1} - 0.3Y_{t-2} + (0.8 + 0.5Y_{t-1})\epsilon_{t-1} + \epsilon_t$

NL-6: $Y_t = 0.5 - (0.4 - 0.4\epsilon_{t-1})Y_{t-1} + \epsilon_t$

NL-7: $Y_t = 0.8\epsilon^2_{t-2} + \epsilon_t$

NL-8: $Y_t = \epsilon_t + 0.3\epsilon_{t-1} + (0.2 + 0.4\epsilon_{t-1} - 0.25\epsilon_{t-2})\epsilon_{t-2}$,

where $\{\epsilon_t\}$ in each model is an independent $N(0,1)$ sequence. The first four models were proposed by Keenan (1985) and also analyzed by Peña and Rodríguez (2002, 2006), whereas the next four models were analyzed by Psaradakis and Vávra (2019). For each model, 1000 replicates of sample size $n = 204$ are generated and an $AR(p)$ model is fitted to the data, where p is selected by the Akaike information Criterion (AIC) with $p \in \{1, 2, 3, 4\}$.

Table 3.5: Powers of the tests at 5% for eight nonlinear models, AR(p) model fitted, n=204

m	Model	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}^*_W$	Q^*_W	$\tilde{M}^*_W$	M^*_W	$\tilde{D}^*$	D^*
7	NL-1	0.136	0.124	0.138	0.130	0.152	0.145	**0.159**	0.151	0.158	0.148
	NL-2	0.517	0.498	0.519	0.506	**0.613**	0.604	0.600	0.590	0.595	0.585
	NL-3	0.936	0.930	0.931	0.925	**0.967**	0.962	0.966	0.962	0.966	0.960
	NL-4	0.866	0.851	0.850	0.834	**0.925**	0.921	0.916	0.910	0.916	0.909
	NL-5	0.850	0.833	0.836	0.822	**0.916**	0.907	0.905	0.894	0.903	0.891
	NL-6	0.871	0.849	0.868	0.846	**0.931**	0.925	0.928	0.923	0.928	0.922
	NL-7	0.390	0.360	0.395	0.386	0.495	0.467	0.499	0.475	**0.500**	0.468
	NL-8	0.448	0.426	0.440	0.418	**0.537**	0.520	0.534	0.517	0.531	0.512
12	NL-1	0.116	0.107	0.110	0.101	**0.154**	0.144	0.150	0.141	0.148	0.138
	NL-2	0.423	0.389	0.412	0.387	**0.548**	0.528	0.537	0.515	0.533	0.504
	NL-3	0.894	0.862	0.886	0.848	**0.957**	0.950	0.947	0.942	0.949	0.941
	NL-4	0.815	0.787	0.783	0.749	**0.886**	0.883	0.876	0.864	0.877	0.865
	NL-5	0.821	0.789	0.783	0.749	**0.901**	0.890	0.891	0.873	0.893	0.873
	NL-6	0.772	0.730	0.789	0.745	0.879	0.865	0.884	0.876	**0.886**	0.874
	NL-7	0.322	0.296	0.332	0.300	**0.442**	0.426	0.440	0.426	0.441	0.420
	NL-8	0.377	0.348	0.373	0.345	**0.518**	0.499	0.505	0.487	0.509	0.479
24	NL-1	0.103	0.089	0.097	0.080	**0.145**	0.129	0.138	0.130	0.133	0.117
	NL-2	0.362	0.326	0.330	0.287	**0.499**	0.473	0.480	0.455	0.466	0.429
	NL-3	0.798	0.755	0.764	0.706	**0.907**	0.891	0.893	0.874	0.895	0.865
	NL-4	0.705	0.656	0.652	0.585	**0.833**	0.814	0.810	0.784	0.810	0.779
	NL-5	0.720	0.674	0.664	0.605	**0.836**	0.811	0.804	0.787	0.806	0.779
	NL-6	0.694	0.648	0.697	0.645	0.841	0.812	**0.845**	0.823	0.837	0.811
	NL-7	0.246	0.203	0.221	0.186	0.356	0.320	**0.360**	0.328	0.348	0.310
	NL-8	0.303	0.271	0.276	0.241	**0.446**	0.418	0.434	0.409	0.416	0.390

The results in Table 3.5 show that the statistics based on the proposed penalized (partial) autocorrelation estimator are more powerful for all nonlinear models and for all values of m. The penalized weighted Ljung-Box test appears to be the most powerful in detecting a nonlinear process, with a few exceptions in which the penalized weighted Monti test or penalized Mahdi-McLeod test is the most powerful.

It's also interesting to check the behavior of the proposed statistics in the detection of nonlinearity in GARCH models. Table 6 and 7 show the power of these tests for four covariance stationary GARCH(1,1) models,

$$Y_t = \varepsilon_t \sigma_t$$

$$\sigma^2 = 1 + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2,$$

where $\alpha \geq 0$, $\beta \geq 0$, and $\alpha + \beta < 1$, and $\{\varepsilon_t\}$ is an independent $N(0,1)$ sequence. Consider the simulation settings from Peña and Rodríguez (2002). Statistics based on the squared residuals are computed, and the empirical power from 1000 replications is found at lags $m = 12$ and $m = 24$. The first two GARCH(1,1) processes considered in Table 3.6 are generated with two sets of parameters GF-1, GF-2 which are from financial time series (Carnero et al., 2001). Models in Table 3.7 are taken from environmental data (Tol, 1996), where the two sets of parameters are GM-1 and GM-2.

$$\text{GF-1: } (\omega, \alpha, \beta) = (1, 0.05, 0.90) \quad \text{GF-2: } (\omega, \alpha, \beta) = (1, 0.15, 0.80)$$

$$\text{GM-1: } (\omega, \alpha, \beta) = (1.21, 0.404, 0.153) \quad \text{GM-2: } (\omega, \alpha, \beta) = (1.58, 0.55, 0.105).$$

Table 3.6: Powers of the tests at 5% for detecting GARCH(1,1) models of financial time series

n	m	Model	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}_W^*$	Q_W^*	$\tilde{M}_W^*$	M_W^*	$\tilde{D}^*$	D^*
250	12	GF-1	0.341	0.324	0.284	0.268	**0.347**	0.335	0.313	0.303	0.309	0.294
		GF-2	0.823	0.822	0.790	0.777	**0.846**	0.842	0.823	0.813	0.818	0.807
	24	GF-1	0.303	0.294	0.238	0.229	**0.339**	0.336	0.287	0.275	0.270	0.252
		GF-2	0.808	0.797	0.725	0.707	**0.857**	0.849	0.810	0.804	0.805	0.791
500	12	GF-1	0.544	0.538	0.480	0.474	**0.566**	0.562	0.519	0.506	0.516	0.503
		GF-2	0.987	0.986	0.980	0.979	**0.992**	**0.992**	0.988	0.988	0.988	0.987
	24	GF-1	0.520	0.514	0.444	0.432	**0.589**	0.584	0.529	0.521	0.518	0.509
		GF-2	0.973	0.972	0.950	0.947	**0.989**	**0.989**	0.979	0.979	0.979	0.977

Table 3.7: Powers of the tests at 5% for detecting GARCH(1,1) models of meteorological time series

n	m	Model	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}^*_W$	Q^*_W	$\tilde{M}^*_W$	M^*_W	$\tilde{D}^*$	D^*
90	12	GM-1	0.394	0.350	0.346	0.306	**0.494**	0.466	0.478	0.449	0.476	0.441
		GM-2	0.553	0.505	0.518	0.461	**0.669**	0.642	0.660	0.631	0.651	0.626
	24	GM-1	0.346	0.301	0.281	0.225	**0.467**	0.438	0.425	0.401	0.404	0.364
		GM-2	0.445	0.385	0.382	0.327	**0.568**	0.535	0.539	0.500	0.530	0.470
180	12	GM-1	0.754	0.729	0.742	0.706	**0.833**	0.818	0.828	0.815	0.829	0.811
		GM-2	0.850	0.829	0.825	0.800	**0.905**	0.898	0.895	0.889	0.896	0.887
	24	GM-1	0.657	0.619	0.606	0.547	**0.772**	0.752	0.744	0.716	0.738	0.702
		GM-2	0.780	0.753	0.739	0.680	**0.867**	0.858	0.845	0.827	0.846	0.820

In general the results show that the proposed statistics are more powerful or tie the unpenalized statistics. The penalized weighted Ljung-Box test performs the best for all cases in Table 3.6 and Table 3.7.

Table 3.8 provides the power for 1000 replicates from a long memory stochastic volatility (LMSV) model (Pérez and Ruiz, 2003),

$$Y_t = exp(h_t/2)\epsilon_t$$

$$(1-\phi B)(1-B)^d h_t = \eta_t$$

where $|\phi| < 1$, and $\{\epsilon_t\}$ and $\{\eta\}$ are assumed to be mutually independent and normally distributed sequences with variances 1 and σ_η^2, respectively. The simulation study follows the design of Fisher and Gallagher (2012). The log volatility follows an AR fractionally integrated MA model (ARFIMA(p, d, q)). The power of the statistic is based on the squared residuals after 1000 replicates of each of the two sample sizes were calculated and tests performed at lags $m = 10$ and $m = 50$. The parameters in six different LMSV models considered are LM-1: $(\phi, d, \sigma_\eta^2) = (0.98, 0, 0.05)$, LM-2: $(\phi, d, \sigma_\eta^2) = (0.9, 0.2, 0.01)$, LM-3: $(\phi, d, \sigma_\eta^2) = (0.9, 0.2, 0.1)$, LM-4: $(\phi, d, \sigma_\eta^2) = (0.8, 0.45, 0.01)$, LM-5: $(\phi, d, \sigma_\eta^2) = (0.9, 0.45, 0.01)$ and LM-6: $(\phi, d, \sigma_\eta^2) = (0, 0.45, 0.1)$. Table 3.8 shows that the proposed tests are more powerful than tests based on sample ACF/PACF for $m = 10$ and $m = 50$ and for all LMSV models. The proposed statistic $\tilde{Q}^*_W$ is almost always the most powerful.

Table 3.8: Powers of the tests at 5% for long memory stochastic volatility processes

n	m	Model	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}^*_W$	Q^*_W	$\tilde{M}^*_W$	M^*_W	$\tilde{D}^*$	D^*
256	10	LM-1	**0.894**	0.892	0.847	0.844	**0.894**	0.892	0.867	0.864	0.863	0.861
		LM-2	**0.284**	0.277	0.257	0.244	0.281	0.276	0.268	0.258	0.264	0.245
		LM-3	0.932	0.931	0.902	0.898	**0.949**	0.947	0.933	0.931	0.932	0.929
		LM-4	0.440	0.426	0.387	0.375	**0.446**	0.432	0.412	0.401	0.408	0.398
		LM-5	**0.835**	0.830	0.784	0.773	0.829	0.828	0.795	0.793	0.792	0.788
		LM-6	0.175	0.164	0.164	0.155	**0.190**	0.179	0.180	0.167	0.176	0.159
	50	LM-1	0.823	0.814	0.703	0.683	**0.876**	0.872	0.835	0.826	0.811	0.794
		LM-2	0.207	0.199	0.135	0.123	**0.273**	0.261	0.219	0.197	0.171	0.152
		LM-3	0.808	0.781	0.692	0.666	**0.891**	0.887	0.843	0.829	0.831	0.814
		LM-4	0.342	0.337	0.250	0.232	**0.425**	0.406	0.356	0.336	0.304	0.283
		LM-5	0.709	0.696	0.588	0.566	**0.807**	0.798	0.736	0.720	0.690	0.674
		LM-6	0.137	0.131	0.103	0.092	**0.164**	0.158	0.150	0.139	0.122	0.106
512	10	LM-1	0.996	0.996	0.995	0.995	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**	**0.998**
		LM-2	0.477	0.471	0.432	0.421	**0.483**	0.473	0.455	0.443	0.450	0.440
		LM-3	**0.995**	**0.995**	0.993	0.993	**0.995**	**0.995**	0.994	0.994	0.994	0.994
		LM-4	0.717	0.712	0.660	0.653	**0.721**	0.719	0.686	0.681	0.686	0.678
		LM-5	0.990	0.990	0.976	0.975	**0.992**	0.989	0.975	0.975	0.975	0.975
		LM-6	0.310	0.304	0.277	0.269	**0.319**	0.312	0.305	0.294	0.303	0.290
	50	LM-1	0.987	0.987	0.968	0.965	**0.993**	**0.993**	0.989	0.989	0.989	0.989
		LM-2	0.387	0.381	0.307	0.296	**0.479**	0.473	0.417	0.403	0.391	0.384
		LM-3	0.974	0.969	0.933	0.922	**0.991**	0.990	0.983	0.982	0.983	0.982
		LM-4	0.654	0.652	0.546	0.529	**0.748**	0.744	0.671	0.661	0.650	0.639
		LM-5	0.978	0.976	0.935	0.930	**0.988**	**0.988**	0.975	0.973	0.971	0.967
		LM-6	0.289	0.281	0.219	0.212	**0.342**	0.332	0.294	0.280	0.267	0.259

3.5.3 Fitted ARCH Processes

To analyze the empirical size and power of the penalized (weighted) statistic from Section 3.4.2, 1000 ARCH series with sample sizes $n = 100, 200, 300, 400$ are generated. The simulation experiment design is from Fisher and Gallagher (2012) and Li and Mak (1994).

Table 3.9 shows the empirical Type I errors where the data are generated from two $AR(1)-ARCH(b)$ models. The parameters are ARCH-1: $(\phi, \omega, \alpha) = (0.2, 0.2, 0.2)$ where ARCH-1 has $b = 1$ and ARCH-2: $(\phi, \omega, \alpha_1, \alpha_2) = (0.2, 0.2, 0.2, 0.2)$ with $b = 2$. We can observe that the empirical sizes for the penalized (weighted) Li and Mak statistics are comparable to the ones for (weighted) Li and

Mak statistics, and all statistics are close to or less than the nominal size 0.05.

Table 3.9: Empirical sizes at 5% for ARCH-1 and ARCH-2 models

	m=6				m=12			
n	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$
ARCH-1 model, b=1								
100	0.022	0.021	0.021	0.019	0.026	0.024	0.027	0.025
200	0.044	0.044	0.037	0.036	0.035	0.034	0.043	0.041
300	0.042	0.041	0.035	0.032	0.046	0.042	0.039	0.034
400	0.043	0.04	0.047	0.044	0.046	0.045	0.045	0.040
ARCH-2 model, b=2								
100	0.022	0.018	0.025	0.025	0.025	0.021	0.023	0.021
200	0.044	0.043	0.046	0.046	0.028	0.028	0.034	0.030
300	0.039	0.038	0.039	0.038	0.039	0.038	0.037	0.035
400	0.053	0.053	0.051	0.048	0.038	0.037	0.041	0.040

Table 3.10 shows the power of these tests for 1000 replicates from two ARCH models: ARCH-2 given above and ARCH-3 where the parameters are (Fisher and Gallagher, 2012)

$$(\phi, \omega, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5) = (0.2, 0.2, 0.2, 0.2, 0.1, 0.05, 0.05).$$

For each simulated sample, AR(1)-ARCH(1) and AR(1)-ARCH(2) are fit. The proposed tests are more powerful for detecting the ARCH processes under fit in all considered cases. The penalized weighted Li-Mak statistic is the most powerful.

Table 3.10: Power levels at 5% for ARCH(1) and ARCH(2) models

	m=6				m=12			
n	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$
AR(1)-ARCH(1) fitted to ARCH-2, b=1								
100	0.157	0.148	**0.185**	0.180	0.097	0.089	**0.130**	0.120
200	0.344	0.332	**0.408**	0.400	0.275	0.251	**0.362**	0.345
300	0.498	0.483	**0.598**	0.589	0.433	0.414	**0.529**	0.514
400	0.639	0.630	**0.719**	0.715	0.546	0.530	**0.665**	0.651
AR(1)-ARCH(2) fitted to ARCH-3, b=2								
100	0.089	0.088	**0.1**	0.097	0.075	0.069	**0.088**	0.078
200	0.209	0.202	**0.226**	0.219	0.159	0.149	**0.205**	0.193
300	0.311	0.307	**0.326**	0.321	0.272	0.257	**0.339**	0.334
400	0.438	0.437	**0.451**	0.449	0.341	0.336	**0.424**	0.418

3.6 Applications

In this section, the penalized statistics are employed to two time series. In each case goodness of fit is assessed using the tests discussed in this chapter. P-values are calculated using (3.16).

Consider the weekly egg prices (per egg in Deutsch Marks/100) paid to the producer from April 1967 to May 1990 (Fan and Yao, 2003; Fan, 2004; Finkenstädt, 2012). The observed series and it's first order differences can be seen in Figure 3.4. Fan and Yao (2003) found that ARMA(1,2) model provides an adequate fit. Here, both an ARMA(1,1) model and an ARMA(1,2) model are fit to the sample of first order differences. Table 3.11 provides the p-values of the ten portmanteau statistics applied to the residuals with lags $m = 10, 15, 30$ and $\alpha = 0.05$. Simulations given above indicate that the penalized version of the ARMA goodness-of-fit statistics are more sensitive in detecting model under fit. In Table 3.11 the p-value of each penalized test is slightly smaller than that of it's non-penalized version. Neither Q_W nor $\tilde{Q}_W$ detect the under fit of the ARMA(1,1) at the 0.05 level, using any of the considered m. In nearly all cases, the penalized and unpenalized version of the statistic indicate the same conclusion. However, using $m = 30$, the model adequacy

is rejected using $\tilde{M}_W$, but not rejected using M_W (boldface in the table).

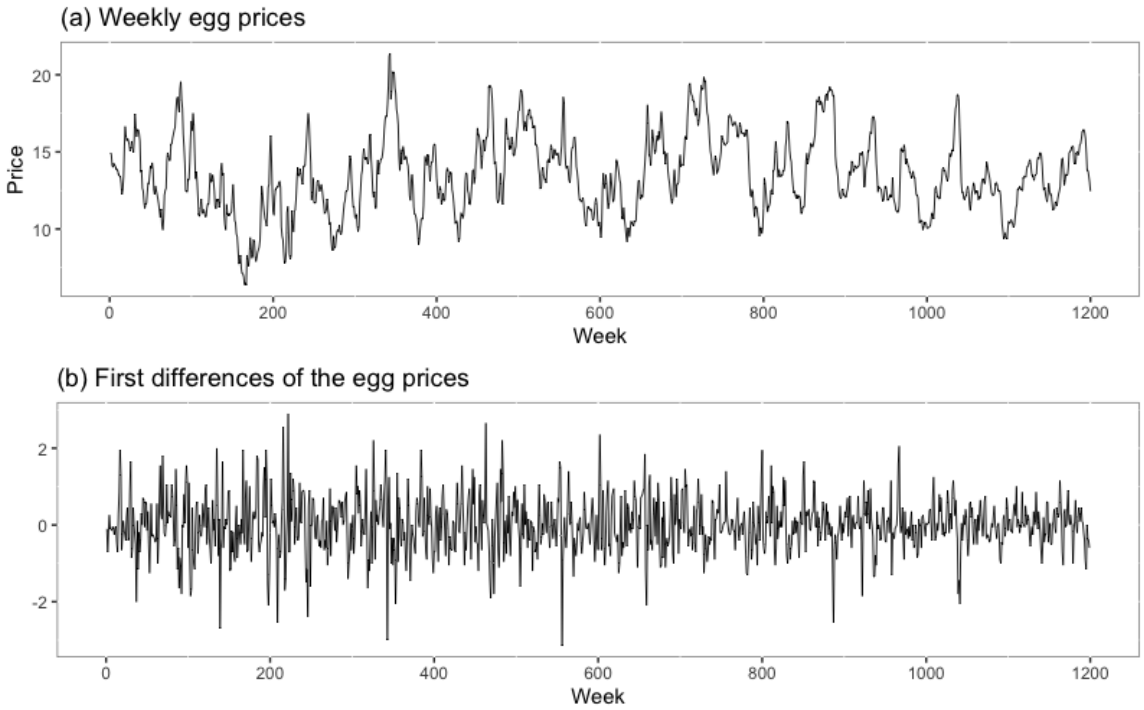


Figure 3.4: (a) Weekly egg prices (per egg in Deutsch Marks/100) paid to the producer from April 1967 to May 1990. (b) First differences of the prices.

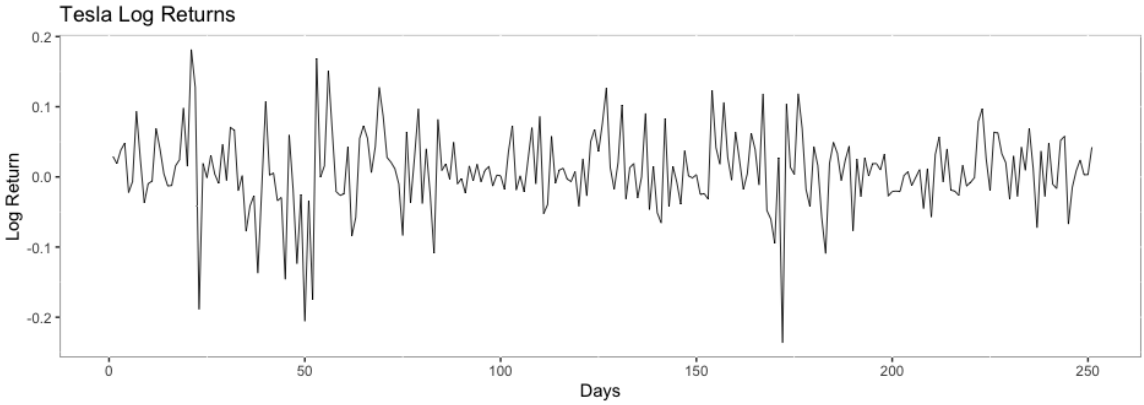


Figure 3.5: Tesla log returns from January 1, 2020 to December 31, 2020

Table 3.11: The p-values of statistics after ARMA models have been fit to Egg price series.

m	$\tilde{Q}$	Q	$\tilde{M}$	M	$\tilde{Q}_W$	Q_W	$\tilde{M}_W$	M_W	$\tilde{D}$	D
					ARMA(1,1)					
10	0.0106	0.0120	0.0068	0.0076	0.0727	0.0763	0.0563	0.0594	0.0454	0.0478
15	0.0793	0.0888	0.0416	0.0475	0.0609	0.0667	0.0382	0.0424	0.0305	0.0345
30	0.0109	0.0126	0.0008	0.0009	0.1006	0.1169	**0.0427**	**0.0511**	0.0341	0.0403
					ARMA(1,2)					
10	0.0984	0.1002	0.1033	0.1051	0.5906	0.5838	0.5997	0.5924	0.1903	0.1898
15	0.3108	0.3154	0.2641	0.2681	0.4942	0.4932	0.4680	0.4672	0.2112	0.2143
30	0.0849	0.0854	0.0859	0.0856	0.571	0.5837	0.5624	0.5743	0.373	0.3794

Consider the log of the daily closing return of Tesla stock (TSLA) on the market days from Jan 1, 2020, through December 31, 2020. The series is plotted in Figure 3.5.

Table 3.12 presents the p-values of ten statistics using squared returns at various lags m, with nearly all tests indicating the presence of nonlinear effects. Once again, in the case of an apparently under fit model, the p-values for the penalized tests are smaller than those of the standard tests, and in nearly all cases the two versions of the test provide the same conclusion. However, using a type I error bound of 0.05 and $m = 40$, the test statistic M^* fails to detect the nonlinearity, while the statistic $\tilde{M}^*$ indicates statistically significant nonlinear effects; similarly, when $m = 60$, M_W^* and $\tilde{M}_W^*$ give different conclusions.

Table 3.13 contains p-values of Li-Mak and weighted Li-Mak statistics for detecting ARCH under fit, with nearly all tests rejecting the adequacy of the ARCH(1) model. All of the tests fail to reject the goodness of fit of the ARCH(2). Note that in all cases the penalized version of the Li-Mak test has higher p-value for ARCH(2) fit than the ordinary Li-Mak test, and the same holds for several choices of m in the weighted version. Although, the p-value of the penalized statistics is typically smaller when the model is under fit, when the fit is adequate, the penalized test can have a *larger* p-value, thus providing more confidence in the adequacy of the fitted model.

Table 3.12: The p-values in detecting nonlinear effects in Tesla returns using squared values

m	$\tilde{Q}^*$	Q^*	$\tilde{M}^*$	M^*	$\tilde{Q}^*_W$	Q^*_W	$\tilde{M}^*_W$	M^*_W	$\tilde{D}^*$	D^*
10	0.0007	0.0013	0.0082	0.0138	0.0000	0.0001	0.0006	0.0014	0.0005	0.0013
20	0.0068	0.0117	0.0632	0.0876	0.0002	0.0004	0.0054	0.0104	0.0046	0.0108
30	0.0010	0.0017	0.0325	0.0431	0.0005	0.0012	0.0153	0.0245	0.0157	0.0288
40	0.0015	0.0023	**0.0402**	**0.0516**	0.0003	0.0007	0.0140	0.0224	0.0192	0.0326
50	0.0056	0.0084	0.1295	0.1528	0.0004	0.0009	0.0199	0.0293	0.0322	0.0492
60	0.0271	0.0367	0.3383	0.3630	0.0008	0.0014	**0.0397**	**0.0506**	0.0668	0.0899

Table 3.13: The p-values of statistics after ARCH Models have been fit to Tesla returns

m	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$	$\tilde{L}(b,m)$	$L(b,m)$	$\tilde{L}_W(b,m)$	$L_W(b,m)$
	ARCH(1)				ARCH(2)			
10	0.0015	0.0031	0.0001	0.0002	0.1285	0.1282	0.1272	0.1289
20	0.0153	0.0264	0.0006	0.0014	0.1907	0.1883	0.1475	0.1513
30	0.0095	0.0151	0.0021	0.0045	0.3876	0.3698	0.2256	0.2283
40	0.0165	0.0244	0.0028	0.0049	0.5927	0.5677	0.3286	0.3198
50	0.0556	0.0735	0.0048	0.0081	0.7021	0.6681	0.4149	0.3979
60	0.1871	0.2211	0.0108	0.0161	0.8236	0.7933	0.5325	0.4998

3.7 Summary

The general forms of test statistics given in (3.8), (3.21) and (3.26), capture most of the tests proposed in the literature. Ideally, the power of these tests could be maximized with optimal choices for weighting sequence $\{\omega_k\}$ and function f. It is fairly easy to see that there can be no power maximizing choice of weights or power maximizing choice of f, since the power depends on the model under the alternative hypothesis. Gallagher and Fisher (2015) consider general weighting choices and Mahdi (2016) considers tests for seasonal ARMA models, while Fisher and Robbins (2018) show that the choice $f(x) = -log(1-x)$ can improve power over $f(x) = x$. While there can be no power optimal choices for target and tuning parameters, the test specific choices for target and tuning parameters given in (3.9) and (3.10), respectively, improve power for each chosen Portmanteau test

statistic. In theory the target and tuning parameters could adapt to the specific data under study if these were selected via cross validation (CV). However, special care must be taken when cross validating dependent data (Burman et al., 1994; Arlot and Celisse, 2010). Adapting these methods to select tuning parameters in the goodness of fit framework did not provide improvement over the proposed plug-in choices provided above.

The general penalized M-estimator (1.12) can potentially be used to improve other time series methods. Formulating the estimation problem in terms of best linear prediction, indicates a natural application of the estimator to forecasting. Banded and tapered estimators stabilize autocorrelation and partial autocorrelation estimators as a function of the lag and can improve forecasting (McMurry and Politis, 2015; Proietti and Giovannelli, 2018). These estimators use the sample ACF at small lags, shrink moderate lag estimates toward zero and set larger lag estimates equal to zero. Thus, the estimators still tend to underestimate the magnitude non-zero correlations. Combining lag-based shrinkage of the banded and tapered estimator with the penalized M-estimator which can increase the magnitude of estimates, has a potential for improving forecasting. Recently, Bergmeir et al. (2018) has demonstrated that K-fold cross validation outperforms traditional out of sample methods for prediction in autoregressive settings. Perhaps their proposed methods can be used to select parameters in (1.12) to improve autoregressive forecasting. A long standing problem in time series econometrics is accurate estimation of standard errors when conducting inference with autocorrelated data (Berk, 1974; Newey and West, 1987). In many cases the goal is to make inference in terms of regression parameters without assuming a specific model for the correlation structure. The popular heteroscedasticity and autocorrelation consistent (HAC) methods (Andrews, 1991), estimate the standard error of the estimated mean (regression) parameters using a weighted linear combination of estimated autocorrelations, or a spectral estimate which can be expressed as such a combination. These methods can result in inflated type I error due to the bias of the correlation estimates, especially in small samples with high amounts of positive correlation (Müller, 2014). The penalized M-estimators of autocorrelation could potentially improve performance of HAC methods. These future directions will be explored elsewhere.

Chapter 4

Prediction Using Penalized ACF/PACF

4.1 Introduction

In prediction theory and time series analysis, the autocorrelation matrix of a stationary process plays a key role. The n-dimensional autocorrelation matrix is a Toeplitz matrix that depends solely on the autocorrelation from lag 0 to $n-1$ for stationary process. Toeplitz correlation matrices are special cases with distinct correlations, one for each diagonal away from the main diagonal, which are used in optimal linear prediction and signal processing. However, the sample autocovaraince or autocorrelation matrix is not consistent when the dimension is the same as the sample size (Wu and Pourahmadi, 2003, 2009). Back to the 1920's, the idea of banding a stationary covariance matrix or limiting moving average MA and autoregressive AR model fitting was investigated by Yule and Slutsky. Banding sample covariance matrix and thresholding have been proposed by Wu and Pourahmadi (2003, 2009) and Bickel and Levina (2008) in longitudinal or time dependent data analysis. If we wish to use the most recent p observations to predict future values, we can fit an autoregressive AR(p) model to the observed data; to choose the order p from the data we can use the Akaike information criterion (AIC). Bickel and Gel (2011) proposed a banded regularization of autocovariance matrices which can fit a longer AR(p) model,and also control the number of parameters to be estimated precisely and the level of accuracy. McMurry and Politis (2010) proposed

a tapered and banded estimator by leaving the main diagonals of the sample autocovariance matrix intact while gradually shrink off-diagonal entries toward zero. They developed the theory of best linear prediction using the resulting autocovariance sequences (McMurry and Politis, 2015) and investigated the asymptotic properties of the tapered and banded estimator. Proietti and Giovannelli (2018) introduced an alternative estimator which is consistent and positive-definite estimator of the autocovariance matrix by using a modified Durbin-Levinson algorithm with tapered and banded sample partial autocorrelation. However, the sample (partial) autocorrelation underestimates the magnitude of the low lag correlations. In this chapter, we discuss the application of the penalized ACF and PACF in prediction based on the methods of McMurry and Politis (2015) and Proietti and Giovannelli (2018).

In order to alleviate the bias and reduce MSE of sample ACF/PACF and improve the prediction for stationary processes, we propose penalized estimators of ACF and PACF where the penalty is based on the distance between correlation and a target. The general penalized objective function is in (1.12). The target values and tuning parameters can depend on the prediction application and the data. This chapter considers using correlation estimators by minimizing (1.12) to reduce the prediction errors for stationary processes. By selecting target and tuning parameters in (1.12), one can increase the moderate and large magnitude correlations toward unity and shrink the small magnitude correlations toward zero to reduce the bias of the correlation estimators and the prediction errors. The penalized ACF and PACF are used for prediction in two ways. First, we propose a penalized autocorrelation matrix estimator and impose positive definiteness, so that the tapered and banded method can be used to construct a penalized predictor. Second, we use a regularized Durbin-Levionson algorithm (Proietti and Giovannelli, 2018; Durbin, 1960) based on tapered and banded sample partial autocorrelation sequence for prediction.

The remainder of this chapter is structured as follows. Section 2 presents our proposed PACF/ACF estimator applied to linear prediction. Section 3 provides a comparative study of the prediction errors for ARMA models, and data application is given in Section 4. Section 5 provides discussion.

4.2 Method

4.2.1 Penalized Estimator

Recall that the general form of penalized correlation estimator is a convex combination of the unpenalized estimator and the target,

$$\tilde{\rho}(h) = w_h \rho_\tau(h) + (1 - w_h)\hat{\rho}(h)$$

where $w_h = \frac{\lambda_h}{1+\lambda_h}$. It has been shown that the penalized estimator follows the same asymptotic distribution as the sample ACF $\hat{\rho}$. To reduce the estimation bias and prediction error, we suggest to select the target as follows,

- if $\hat{\rho}(h)$ less than one standard error from 0, $|\rho_\tau(h)| < |\hat{\rho}(h)|$;
- if $\hat{\rho}(h)$ is larger, then ρ_τ is larger than $\hat{\rho}(h)$ with the increase depending on the sample size and the magnitude of $\hat{\rho}(h)$.

The choice of target and tuning parameters satisfying the principles above are given as,

$$\rho_\tau(h) = \begin{cases} 0 & |\hat{\rho}(h)| \leq \ell_h \\ min\{\hat{\rho}(h) + \sqrt{\frac{1}{n}} v_h, 1\} & |\hat{\rho}(h)| > \ell_h \end{cases} \tag{4.1}$$

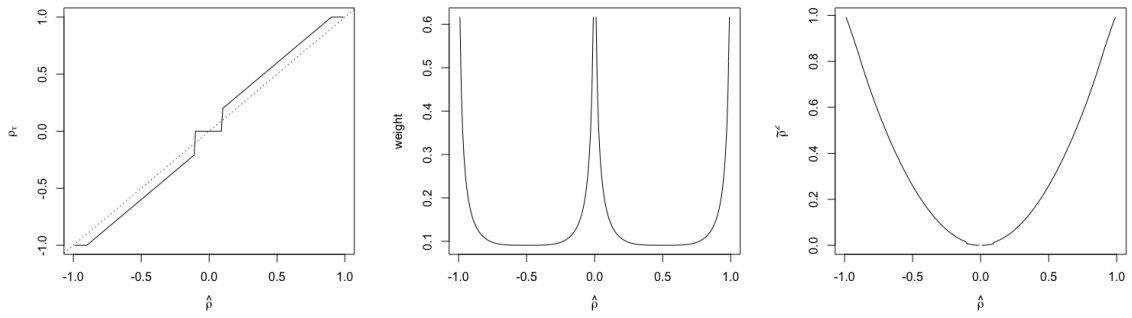
and

$$\lambda_h = \frac{\ell_h}{sin(|\hat{\rho}|(1 - |\hat{\rho}|) * 2\pi)}$$

then the penalized estimator is,

$$\tilde{\rho}(h) = \frac{\hat{\rho}(h) + \lambda_h sgn(\hat{\rho}(h))|\rho_\tau(h)|}{1 + \lambda_h} \tag{4.2}$$

where $\ell_h = \sqrt{(n-h)/(n^2+2n)}$, $\hat{\rho}(h)$ is the lag h sample (PACF) ACF of the residual sequence, and $v_h = \frac{1.08}{1+exp((h-25))/10}$ is a lag decay. Figure (4.1) demonstrates the target, weight and resulting penalized estimator for $\hat{r}(1)$, when $h = 1$ and $n = 100$. Part (a) shows how ρ_τ depends on $\hat{\rho}$. The weight function is graphed as a function of $\hat{\rho}$ in part (b). Finally, part (c) of the figure shows the map from $\hat{\rho}(1)$ to $\tilde{\rho}^2(1)$.



(a) Target (b) Weight (c) Penalized Estimator

Figure 4.1: (a) ρ_τ, (b) w_h, and (c) $\tilde{\rho}^2$ as functions of $\hat{\rho} = \hat{r}(1)$, $n = 100$, and $h = 1$.

As in Section 2.2, it is no guarantee that the penalized estimator is non-negative definite. Since the inverse of $\hat{\boldsymbol{R}}$ is a important element in prediction, the autocorrelation matrix can be corrected to achieve finite-sample positive definiteness. We apply one way to modify the estimated function to be non-negative definite as below. More techniques to impose positively are discussed in McMurry and Politis (2010). Given the sample autocorrelation matrix and the penalized ACF matrix,

$$\hat{\boldsymbol{R}}_h = \begin{pmatrix} 1 & \hat{\rho}(1) & ... & \hat{\rho}(h-1) \\ \hat{\rho}(1) & 1 & ... & \hat{\rho}(h-2) \\ \cdots & \cdots & ... & \cdots \\ \hat{\rho}(h-1) & \hat{\rho}(h-2) & ... & 1 \end{pmatrix} \quad \tilde{\boldsymbol{R}}_h = \begin{pmatrix} 1 & \tilde{\rho}(1) & ... & \tilde{\rho}(h-1) \\ \tilde{\rho}(1) & 1 & ... & \tilde{\rho}(h-2) \\ \cdots & \cdots & ... & \cdots \\ \tilde{\rho}(h-1) & \tilde{\rho}(h-2) & ... & 1 \end{pmatrix}$$

with the smallest eigenvalues $\alpha = min(\text{eigenvalue of } \hat{\boldsymbol{R}}_h)$ and $\beta = min(\text{eigenvalue of } \tilde{\boldsymbol{R}}_h)$. Let $c = |\beta|/(\alpha + |\beta|)$, then

$$\tilde{\rho}^*(h) = \alpha\hat{\rho}(h) + (1-\alpha)\tilde{\rho}(h)$$

and

$$\boldsymbol{R}_h^* = \begin{pmatrix} 1 & \tilde{\rho}^*(1) & ... & \tilde{\rho}^*(h-1) \\ \tilde{\rho}^*(1) & 1 & ... & \tilde{\rho}^*(h-2) \\ \cdots & \cdots & ... & \cdots \\ \tilde{\rho}^*(h-1) & \tilde{\rho}^*(h-2) & ... & 1 \end{pmatrix}.$$

Similarly, the penalized partial autocorrelation estimator can be formulated as

$$\tilde{\alpha}(h) = w_h \alpha_\tau(h) + (1 - w_h)\hat{\alpha}(h) \tag{4.3}$$

where $\hat{\alpha}(h)$ is the ordinary sample PACF, $\alpha_\tau(h)$ is a target value and $w_h = \lambda_h/(1+\lambda_h)$. The target values and tuning parameters can be selected using the sample PACF.

4.2.2 Tapered and Banded Predictor

Consider a realization $\{Y_t, t = 1, ...n\}$ of a stationary random process with mean zero and autocovariance function $\gamma_h = E[Y_t Y_{t-h}]$. To predict Y_{n+1} based on these observed data, the $n \times n$ Toeplitz autocorrelation matrix $\boldsymbol{R}_n$ plays a central role. With respect to mean squared error, the full sample optimal linear predictor (FSO) is

$$\hat{Y}_{n+1} = \hat{\phi}_{n1} Y_n + \hat{\phi}_{n2} Y_{n-1} + ... + \hat{\phi}_{nn} Y_1 \tag{4.4}$$

where the coefficients $\hat{\phi}_{ni}$ are given by $\hat{\boldsymbol{\phi}}_n = [\hat{\phi}_{n1}, ..., \hat{\phi}_{nn}] = \hat{\Gamma}_n^{-1}\hat{\boldsymbol{\gamma}}_{\boldsymbol{n}} = \hat{\boldsymbol{R}}_n^{-1}\hat{\boldsymbol{\rho}}_n$ and $\hat{\Gamma}_n = [\hat{\gamma}_{|i-j|}]_{i,j=1}^n$ is the sample autocovariance matrix of $Y_1, ..., Y_n$, $\hat{\boldsymbol{\gamma}}_n = (\hat{\gamma}_1, ..., \hat{\gamma_n})^{'}$ is the vector of sample covarainces at lags $1, ..., n$. Bickel and Gel (2011) investigated a predictor of Y_{n+1} that use the $p \times p$ submatrix of the banded regularization of sample autocovariance matices with $p = o(n)$. In practice, the coefficient vector $\hat{\boldsymbol{\phi}}_{\boldsymbol{n}} = (\hat{\phi}_{n1}, ..., \hat{\phi}_{nn})^{'}$ is routinely truncated to its first p components in order to be consistently estimated. This procedure is the same as fitting an AR(p) process to the data. The resulting predictor, which is called partial sample optimal predictor (PSO), is

$$\hat{Y}_{n+1}^p = \hat{\phi}_{p1} Y_n + \hat{\phi}_{p2} Y_{n-1} + ... + \hat{\phi}_{pp} Y_{n-p+1} \tag{4.5}$$

where the coefficients $\hat{\phi}_{pi}$ are given by $\hat{\boldsymbol{\phi}}_p = [\hat{\phi}_{p1}, ..., \hat{\phi}_{pp}] = \hat{\Gamma}_p^{-1}\hat{\boldsymbol{\gamma}}_p$ and $\hat{\Gamma}_p^{-1}$ can be corrected to positive definiteness (McMurry and Politis, 2015). For example, the predictor proposed in Bickel and Gel (2011) has coefficients with $p = n^{1/2}$ and the AR prediction chooses by AIC minimization with $p = p_{AIC}$. Note the sample autocovariance matrix $\hat{\Gamma}_n = [\hat{\gamma}(|i-j|)], i, j = 1, ..., n$ is not consistent. Wu and Pourahmadi (2009) and Bickel and Gel (2011) proposed the banded autocovariance estimator at lag h which is

$$\hat{\gamma}_{h,B} = \hat{\gamma}(h) I(h \leq l), \quad 0 \leq h \leq n$$

where $I(.)$ is the indicator function and l is the banding parameter. This estimator is consistent for the class of nonlinear short-range dependent processes and under the Frobenius norm (Bickel and Gel, 2011).

The taper and banded autocovariance estimator at lag h is proposed by McMurry and Politis (2010) estimates lag h covariance as

$$\hat{\gamma}_{h,TB} = \hat{\gamma}(h)\omega(h), \quad 0 \leq h \leq n,$$

with $\omega(h) = \kappa(h/l)$ where l is a banding parameter, and they suggest $\kappa(.)$ to be the trapezoidal taper

$$\kappa(z) = \begin{cases} 1 & \text{if } |z| \leq 1 \\ 2 - |z| & \text{if } 1 < |z| \leq 2 \\ 0 & \text{if } |z| > 2. \end{cases} \tag{4.6}$$

Using the tapered and banded sample autocovariance, McMurry and Politis (2015) developed the corresponding optimal linear predictor. Our goal is to predict Y_{n+1} based on these observed data. Applying the proposed autocorrelation estimator $\tilde{\rho}(h)$ for each lag $h = 0, 1, ..., n-1$, we shall define the estimated ACF using the data-based choice of the banding parameter with corrections towards positive definiteness. So the tapered and banded autocorrelation estimator using penalized method is given as

$$\tilde{\rho}_{h,PTB} = \kappa(h/l)\tilde{\rho}(h) \tag{4.7}$$

Selection of the banding parameter l Both the FSO and PSO predictors of equations (4.4) and (4.5) are affected by the banding parameter l selection. McMurry and Politis (2010) provided a data-based rule for picking l that $\hat{l}$ is the smallest positive integer such that $|\hat{\rho}(l+k)| < c(logn/n)^{1/2}$ for $k = 1, ..., K_n$. The empirical rule for picking l remains valid for all $c > 0$ and $1 \leq K_n \leq n$. Following the suggestion of McMurry and Politis (2010, 2015), we use $c = 2$ and $K_n = 5$. The rule entails performing an approximate 95% test of the null hypothesis $\rho(\hat{l}+1), ..., \rho(\hat{l}+K_n)$, which are all simultaneously equal to zero.

In this work, the proposed autocorrelation estimation $\tilde{\rho}$ in (4.2) is applied at lag $h = 1, ..., p$

instead of the sample ACF. Thus the penalized predictor of Y_{n+1} is defined as

$$\tilde{Y}_{n+1}^{p} = \tilde{\phi}_{p1}(n)Y_n + \tilde{\phi}_{p2}(n)Y_{n-1} + ... + \tilde{\phi}_{pp}(n)Y_{n-p+1} \tag{4.8}$$

where p is the banding parameter and the coefficient vector $\tilde{\boldsymbol{\phi}}_p = [\tilde{\phi}_{p1}(n), ..., \tilde{\phi}_{pp}(n)]'$ is given by the Yule-Walker equations: $\tilde{\boldsymbol{\phi}}_p = (\tilde{\boldsymbol{R}}_p^*)^{-1}\tilde{\boldsymbol{\rho}}_{p,PTB}^*$ where $\tilde{\boldsymbol{R}}_p^*$ and $\tilde{\boldsymbol{\rho}}_{p,PTB}^*$ is the penalized autocorrelation matrix and vector that result after $\tilde{\boldsymbol{R}}_p$ is corrected to positive definiteness. In practice, We can also use penalized PACF to calculate AIC to pick the banding parameter $p = p_{AIC}$. So the banding parameter can be p_{AIC} or $2l$. In this situation, the penalized coefficient vector $\tilde{\boldsymbol{\phi}}_p$ can be calculated by the Durbin-Levinson algorithm using the penalized PACF estimator using equation (4.3).

4.2.3 Regularized Durbin-Levinson Algorithm with Penalized Estimator

Proietti and Giovannelli (2018) introduced the regularized Durbin-Levinson algorithm, which shrinks the sample partial autocorrelation $\hat{\phi}_{hh}$ towards zero with $\hat{\pi}_{hh} = \omega_h\hat{\phi}_{hh}$ where $\omega_h = \kappa(h/l)$ is the trapezoidal kernel given in (4.6). Along with the estimated autoregressive coefficients $\hat{\pi}_{hj}$, the regularized sample autocovariance matrix can be obtained. The regularized Durbin-Levinson autocovariance estimator is consistent and positive definite under some suitable assumptions. In this section, we modify the regularized Durbin-Levinson algorithm using the penalized estimator to proposed a new autocorrelation matrix and partial autocorrelation matrix.

Let $\gamma_r(h)$ be the regularized autocovariance at lag h, $\phi_{hj}, j = 1, ..., h$ be the usual coefficient without regularization, $\pi_{hj}, j = 1, ..., h$ be the coefficient of the regularized predictor, $\tilde{\gamma}_r(h)$ be the regularized autocovariance using the penalized method, $\tilde{\phi}_{hj}, j = 1, ..., h$ be the usual coefficient without regularization, and $\tilde{\pi}_{hj}$ be the penalized coefficient of the regularized predictor, Algorithm 1 displays the procedure of the penalized and regularized Durbin-Levinson algorithm. We obtain the penalized PACF as given in equation (4.3) and $\tilde{\alpha}(h) = \tilde{\phi}_{hh}$. If $\omega_h = 1$ for all h and the sample ACF is used, the algorithm is the usual Durbin-Levinson recursion which leads to the sample PACF. We can obtain the usual coefficient of Yule-Walker predictor based on p lag values when $\omega_h = 1$ for $h \leq p$ and $\omega_h = 0$ for $h > p$. This regularized algorithm with penalized correlations also yields the regularized penalized autocovariance matrix $\tilde{\boldsymbol{\Gamma}}_n = [\tilde{\gamma}_r(|i-j|)], i, j = 1, .., n$.

In order to predict Y_{t+1} using this algorithm, the optimal linear predictor can be estimated

Algorithm 1: Regularized Durbin-Levinson algorithm using penalized estimators

Input: Penalized PACF $\tilde{\alpha}(h)$ and ACF $\tilde{\rho}(h), h = 1, 2, ..n-1$

Output: Penalized coefficients $\tilde{\pi}_{hj}, j = 1, .., h$ and optimal lag h^*

Start:

$$\nu_0 = \hat{\gamma}(0), \qquad \nu_0^r = \gamma_r(0) = \hat{\gamma}(0)$$
$$\tilde{\phi}_{11} = \tilde{\alpha}(1), \quad \tilde{\pi} = \omega_1 \tilde{\phi}_{11}, \quad \tilde{\gamma}_r(1) = \nu_0^r \tilde{\pi}_{11}$$
$$\nu_1 = (1 - \tilde{\phi}_{11}^2)\nu_0, \quad \nu_1^r = (1 - \tilde{\pi}_{11}^2)\nu_0^r$$
$$AICC_1 = nlog(\nu_1^r) + 2$$

while $h = 2, ..., n-1$ **do**

$$\tilde{\phi}_{hh} = \frac{\tilde{\gamma}(h) - \sum_{j=1}^{h-1} \tilde{\phi}_{h-1,j}\tilde{\gamma}(h-j)}{\nu_{h-1}}$$
$$\tilde{\pi}_{hh} = \omega_h \tilde{\phi}_{hh}, \quad \tilde{\gamma} = \sum_{j=1}^{h-1} \tilde{\pi}_{h-1,j}\tilde{\gamma}_r(h-j) + \nu_{h-1}^r \tilde{\pi}_{hh}$$
$$\tilde{\phi}_{hj} = \tilde{\phi}_{h-1,j} - \tilde{\phi}_{hh}\tilde{\phi}_{h-1,h-j}, \quad \tilde{\pi}_{hj} = \tilde{\pi}_{h-1,j} - \tilde{\pi}_{hh}\tilde{\pi}_{h-1,h-j}, \quad j = 1, 2, ..., h-1$$
$$\nu_h = (1 - \tilde{\phi}_{hh}^2)\nu_{h-1}, \qquad \nu_h^r = (1 - \tilde{\pi}_{hh}^2)\nu_{h-1}^r$$
$$AICC_h = nlog(\nu_h^r) + 2h$$

end

The penalized coefficients $\tilde{\boldsymbol{\phi}}_h$ and optimal lag $h^* = \{h : \min\{AICC_1, ..., AICC_p\}\}$

by

$$\tilde{Y}_{n+1}^n = \sum_{j=1}^{n} \tilde{\pi}_{nj} Y_{n+1-j} \tag{4.9}$$

where the coefficients $\tilde{\pi}_{nj}$ are calculated from the regularized Durbin-Levinson algortihm using the penalized correlations. The proposed estimators follow the assumptions in McMurry and Politis (2015) and Proietti and Giovannelli (2018).

4.3 Simulations

We conducted simulation experiments to access the performance of our proposed estimator in prediction and to compare with the methods in the literature (McMurry and Politis, 2015; Proietti and Giovannelli, 2018). In each case, 1000 replications were performed where an AR(1) or MA(1) model of sample size $n = 101$ was generated using several ϕ or θ parameter values. We used the first n observations to predict the next observation ($n + 1$'st observation) for each simulated time

series. Seven approaches are examined in each prediction. For the tapered and banded method, we selected three predictors based on the sample ACF from McMurry and Politis (2015) and an usual AR predictor to compare with our proposed predictor which applied the tapered and banded technique and penalized correlation estimators. Three tapered and banded predictors are computed by different correction methods towards positive definiteness, see McMurry and Politis (2010, 2015).

AR: the AR predictor with p_{AIC} chosen by AIC minimization.

Pen-PD: the predictor with the penalized ACF/PACF and banded with lag selected via AIC.

TB-WN: the tapered and banded predictor with shrinkage towards white noise.

TB-2o: the tapered and banded predictor with shrinkage towards a 2nd order estimate.

TB-Th: the tapered and banded predictor with threshold correction to positive definiteness.

Using the regularized Durbin Levinson algorithm, the performance of our proposed predictor which is given in 4.9 against the existing predictor in Proietti and Giovannelli (2018) are examined.

Reg-DL: the predictor from Durbin-Levinson regularized algorithm.

Pen-Reg-DL: the predictor using D-L regularized algorithm with penalized ACF/PACF.

For all simulations, accuracy of the prediction is described by root mean squared prediction error (RMSPE). For the predictor Pen-Reg-DL, additional way to examine the forecast ability is applied: $Ratio = \frac{RMSPE_{Reg-DL}}{RMSPE_{Pen-Reg-DL}}$. If $Ratio > 1$, it means that Pen-Reg-DL performs better than Reg-DL for one replicate. We count the number of $Ratio > 1$ to see the proportion that penalized predictor outperforms in each simulation.

4.3.1 AR(1) prediction

In this experiment, we simulated time series of length 101 and used the first 100 values to predict the 101'st. Table 4.1 provides the root mean squared prediction errors for AR(1) processes. Since it directly fits an AR model, this simulation should favor the AR predictor (McMurry and Politis, 2015). We see in Table 4.1 the proposed predictor is better or competitive than the AR and other tapered and banded predictors for most of the cases. For lags greater than 2l, the banding of sample autocorrelation sequences vanish, affecting an MA model. As a result, the AR predictor

will outperform a method that uses a very high order MA model to approximate the low order AR model. The result shows that AR and Pen-PD predictors have smaller RMSPE for most of parameters, except small values of the AR coefficient. This is similar as the simulation results in McMurry and Politis (2015). The smallest RMSPE for each setting is bold.

Table 4.1: RMSPE of (penalized) tapered and banded estimators for $AR(1)$ models

	Root mean square prediction errors				
ϕ	AR	Pen-PD	TB-WN	TB-2o	TB-Th
-0.95	0.9898	**0.9762**	1.4117	1.1460	1.3279
-0.9	1.0069	**1.0040**	1.1374	1.0560	1.1167
-0.7	0.9970	**0.9956**	1.0428	1.0300	1.0377
-0.5	1.0165	**1.0162**	1.0328	1.0431	1.0495
-0.3	1.0303	**1.0273**	1.0453	1.0503	1.0497
-0.1	1.0072	1.0060	1.0034	**1.0033**	1.0035
0.1	1.0143	1.0141	**1.0053**	**1.0053**	**1.0053**
0.3	0.9775	**0.9752**	0.9884	0.9901	0.9902
0.5	1.0445	**1.0346**	1.0729	1.0867	1.0899
0.7	1.0124	**1.0082**	1.0570	1.0473	1.0490
0.9	1.0023	**1.0006**	1.1163	1.0435	1.0966
0.95	1.0597	**1.0487**	1.2963	1.1543	1.2368

Root mean squared prediction errors for the predictors using the regularized Durbin-Levinson algorithm are shown in Table 4.2. One can see that our proposed predictor Pen-Reg-DL performs better than the regularized predictor Pen-DL when the coefficient of AR(1) model is large. The differences in RMSEP between our method and the Reg-DL tend to increase with the magnitude of the AR parameter. In addition, the proportions of $Ratio > 1$ are greater than 50% for almost all the simulation cases.

Table 4.2: RMSPE of the (penalized) Durbin-Levinson regularized estimators for $AR(1)$ models

	Root mean square prediction errors					
ϕ	-0.95	-0.9	-0.7	-0.5	-0.3	-0.1
Reg-DL	1.2228	1.0745	1.0040	1.0121	1.0399	1.0033
Pen-Reg-DL	0.9817	1.0030	0.9964	1.0135	1.0399	1.0036
Pr(Ratio> 1)	0.535	0.506	0.525	0.506	0.492	0.706
ϕ	0.1	0.3	0.5	0.7	0.9	0.95
Reg-DL	1.0038	0.9850	1.0398	1.0185	1.0515	1.1855
Pen-Reg-DL	1.0045	0.9845	1.0382	1.0090	1.0075	1.0559
Pr(Ratio> 1)	0.740	0.502	0.522	0.518	0.524	0.548

4.3.2 MA(1) prediction

Similarly, we simulated time series from MA(1) model of length 101 and predicted the next data point using the first 100 observations. The root mean square prediction errors are provided in Table 4.3. This experiment should be in favor of the tapered and banded predictors since it estimates the correlation of an MA model directly. Note that the banded and tapered predictors is better or competitive with the AR predictor for most of the MA coefficient parameters. Our proposed predictor Pen-PD is better than AR predictor or tapered and banded predictors when the coefficient is small $|\theta| < 0.5$, but not for large θ. It is not surprising since the correlation is moderate when the MA coefficient is large, for example $\rho(1) = 0.49$ when $\theta = 0.9$, the RMSE of the penalized ACF/PACF and sample ACF/PACF is slightly less than one in Table 2.3.

Table 4.4 shows the RMSPE and the proportion of $Ratio > 1$ for the regularized Durbin-Levinson estimators under MA(1) processes. We see in Table 4.4 that the simulated RMSPE of proposed Pen-Reg-DL predictor is comparable to that of Reg-DL. The RMSPE of our proposed predictor is better when $|\theta| < 0.7$. For $|\theta| < 0.7$, The advantage of the proposed predictor is also showed in Table 4.4 in terms of the proportion of Ratio> 1, where Pr(Ratio> 1) is over 50%.

Table 4.3: RMSPE of (penalized) tapered and banded estimators for $MA(1)$ models

	Root mean square prediction errors				
θ	AR	Pen-PD	TB-WN	TB-2o	TB-Th
-0.9	1.0817	1.1021	1.0798	**1.0754**	1.0940
-0.7	1.0725	1.0711	1.0450	**1.0422**	1.0488
-0.5	1.0173	1.0156	**1.0026**	1.0045	1.0059
-0.3	1.0192	**1.0153**	1.0319	1.0332	1.0327
-0.1	1.0089	**1.0018**	1.0062	1.0064	1.0064
0.1	1.0477	1.0467	1.0399	**1.0398**	**1.0398**
0.3	**1.0009**	1.0022	1.0219	1.0225	1.0230
0.5	1.0086	**1.0023**	1.0125	1.0132	1.0159
0.7	1.0492	1.0585	1.0441	**1.0400**	1.0535
0.9	1.0760	1.0878	1.0758	**1.0712**	1.0813

Table 4.4: RMSPE of the (penalized) Durbin-Levinson regularized estimators for $MA(1)$ models

	Root mean square prediction errors				
θ	-0.9	-0.7	-0.5	-0.3	-0.1
Reg-DL	1.1588	1.1019	1.0227	1.0358	1.0050
Pen-Reg-DL	1.1620	1.1021	1.0217	1.0354	1.0054
Pr(Ratio> 1)	0.471	0.491	0.481	0.504	0.715
θ	0.1	0.3	0.5	0.7	0.9
Reg-DL	1.0402	1.0206	1.0381	1.0928	1.1575
Pen-Reg-DL	1.0402	1.0229	1.0375	1.0933	1.1568
Pr(Ratio> 1)	0.748	0.502	0.511	0.470	0.494

4.4 Data Application

To demonstrate the performance of the proposed methods in prediction, consider the analysis of a data example. Following the application analyzed by Bickel and Gel (2011) and Proietti and

Giovannelli (2018), we examined the sea surface temperatures in the Pacific region called Niño 3.4. The monthly time series for the period from Jan. 1950 to Dec. 2019 ($N = 840$), are available from the National Oceanic and Atmospheric Administration (NOAA). The series can be seen in Figure 4.2.

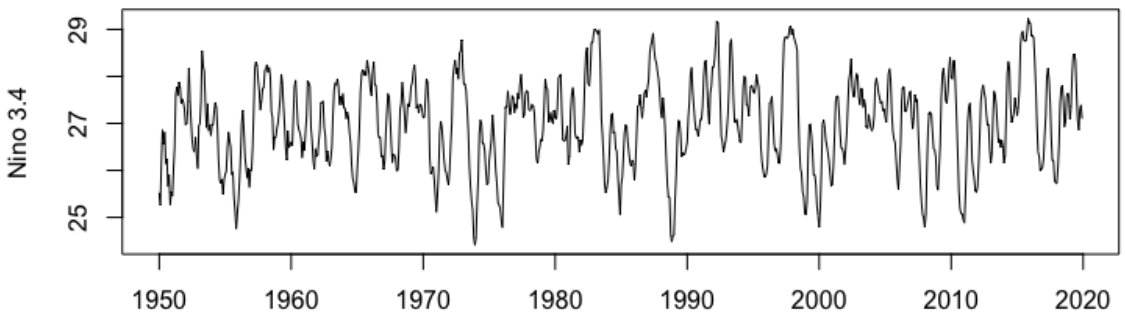


Figure 4.2: El Niño 3.4 from January 1950, to December 2016

A rolling one-step ahead prediction exercise is used to compare the forecasting ability of our proposed predictors using penalized ACF/PACF to the traditional AR predictors, some of the McMurry and Politis predictors, and the regularized Durbin-Levinson estimator (Proietti and Giovannelli, 2018). We compute the one-step-ahead predictors using the training sample of size $n_{tr} = 600$, starting from January 2000. To calculate the root mean square prediction errors, We continue by adding one more future data point and removing the initial one until we reach the end of the sample. Note the number of observations in each prediction is fixed. Then the coefficients of the linear predictor for each forecast round are estimated. The experiment yields $n_t = N - n_{tr} = 240$ prediction errors for each prediction method. Table 4.5 provides the RMSPE for all predictors we discussed in Section 4.3. The results show that the proposed predictor Pen-PD outperforms the tapered and banded predictors (TB-WN, TB-2o, and TB-Th) and the regularized Durbin-Levinson predictor (Reg-DL). Also, the penalized predictor Pen-Reg-DL performs better than Reg-DL in terms of RMSPE. Here Pr(Ratio$>$ 1) is based on the total number of forecast rounds $n_t = 240$. Over 50% of the forecasts the Pen-Reg-DL outperforms.

Table 4.5: Root mean square prediction errors for El Niño 3.4 data

RMSPE							
AR	Pen-PD	TB-WN	TB-2o	TB-Th	Reg-DL	Pen-Reg-DL	Pr(Ratio > 1)
0.2440	0.2473	0.3890	0.2683	0.4563	0.2681	0.2678	0.506

4.5 Summary

We have explored the two penalized based predictors based on the tapered and banded estimator and the regularized Durbin-Levinson estimator. The penalized (partial) autocorrelation estimator has some properties compared with sample ACF. It shrinks towards zero for weak correlation and tends to increase the magnitude of ACF/PACF estimator for larger correlation. Target and tuning parameters can be selected to improve the accuracy of prediction . In addition, we introduced a correction towards positive definiteness using a convex combination of sample ACF/PACF and raw penalized ACF/PACF. When the dimension of the matrix is high or same as the number of observations, the taper and banded estimators (McMurry and Politis, 2015) and regularized Durbin-Levinson estimator (Proietti and Giovannelli, 2018) are consistent. Two new predictors are shown to work well with the penalized ACF/PACF in simulations and application, especially for strong correlated series. Tasks for future research include the further investigation of the penalized estimator for time series when the correlation is moderate, the extension to h-step-ahead prediction in application.

Appendices

Appendix A Autocorrelation Function of AR(p), MA(q) and ARMA(p,q) Processes

Recall that the autoregressive moving average (ARMA) model has gained enormous popularity in various research areas, which is given by:

$$X_t = \mu + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2).$$

This appendix summarizes the autocorrelation functions of some typical ARMA processes.

- **AR(1)**

  The autocorrelation function for AR(1) model is $\rho(h) = \phi_1^h$,

- **AR(2)**

  The autocorrelation function for AR(2) model is

$$\rho(0) = 1, \quad \rho(1) = \frac{\phi_1}{1 - \phi_2}$$

$$\rho(h) = \phi_1 \rho(h-1) + \phi_2 \rho(h-2), \quad h = 2, 3, ...$$

- **MA(q)**

  For MA(q),

$$\rho(h) = \begin{cases} \frac{\sum_{t=0}^{q-h} \theta_t \theta_{t+h}}{1+\theta_1^2+...+\theta_q^2} & 1 \le h \le q \\ 0 & h > q. \end{cases}$$

- **ARMA(1,1)**

  For ARMA(1,1),

$$X_t = \mu + \phi X_{t-1} + \theta e_{t-1} + e_t$$

$$\rho(1) = \frac{(\phi + \theta)(1 + \phi\theta)}{1 + 2\phi\theta + \theta^2}$$

  For $h \ge 2$, we have

$$\rho(h) = \phi^{h-1} \rho(1).$$

Appendix B Functions in R Package PenalizedPortTest

Penalized.cor.gof {PenalizedPortTest}

Penalized.cor.gof

Description

The Penalized ACF/PACF estimator for time series goodness of fit

Usage

```
Penalized.cor.gof(
  x,
  lag = 1,
  fitdf = 0,
  type = c("LB", "BP", "M", "WLB", "WBP", "WM", "MM", "LM", "WLM", "LM.pacf",
    "WLM.pacf"),
  alpha = 0.05
)
```

Arguments

`x`	A numeric vector or univariate time series
`lag`	The lag which the statistic will be based on to calculate penalized acf/pacf, default = 1
`fitdf`	The number of degrees of freedom to be subtracted if x is a series of residuals, default = 0
`type`	The type of test which penalized acf/pacf to be used, default = "LB"
`alpha`	The nominal level selected, default = 0.05

Value

An object of class penalized acf/pacf estimation for time series goodness of fit tests

Examples

```
## Not run:
data <- arima.sim(n=100, model=list(ar=0.9))
pen_cor <- Penalized.cor.gof(x = data, lag=10, fitdf=1, type="LB", alpha=0.05)

## End(Not run)
```

Penalized.Box.test

Description

The Penalized (Weighted) Box-Pierce, (Weighted) Ljung-Box, (Weighted) Monti,

Weighted McLeod-Li type test for fitted ARMA and detection of nonlinear Models

Usage

```
Penalized.Box.test(
  x,
  lag = 1,
  type = c("Ljung-Box", "Box-Pierce", "Monti"),
  fitdf = 0,
  squared = FALSE,
  log.squared = FALSE,
  absolute = FALSE,
  weighted = TRUE,
  penalized = TRUE,
  alpha = 0.05
)
```

Arguments

x	The residuals or initial data which is a numeric vector or univariate time series
lag	The lag which the statistic will be based on to calculate penalized acf/pacf, default = 1
type	the type of test, default = "Ljung-Box"
fitdf	The number of degrees of freedom to be subtracted if x is a series of residuals, default = 0
squared	Take the squared value of residuals to detect nonlinear processes, default=FALSE
log.squared	Take the log of the squared residuals to detect nonlinear processes, default=FALSE
absolute	Take the absolute value of the residuals to detect nonlinear processes, default=FALSE
weighted	If FALSE, perform the original Box-Pierce, Ljung-Box or Monti or McLeod Li type if using a transformation, default=TRUE.
penalized	If FALSE, perform the unpenalized tests using the sample ACF/PACF estimates, default=TRUE
alpha	The nominal level selected, default = 0.05

Value

A list with class "htest" for penalized/traditional (Weighted) Box-Pierce,

(Weighted) Ljung-Box, (Weighted) Monti tests containing statistics, parameters, and p value.

Examples

```
## Not run:
data <- arima.sim(n=100, model=list(ar=0.9))
## Fit AR(1) model
p.fit <- 1
q.fit <- 0
res <- arima(data, order=c(p.fit, 0, q.fit))$resid
out_test1 <- Penalized.Box.test(x = res, lag=10, fitdf=1, type="Ljung-Box", weighted=TRUE, penalized=TRUE)
```

Penalized.MahdiMcLeod.test

Description

The Penalized MahdiMcLeod test for fitted ARMA and detection of nonlinear Models

Usage

```
Penalized.MahdiMcLeod.test(
  x,
  lag = 1,
  fitdf = 0,
  squared = FALSE,
  log.squared = FALSE,
  absolute = FALSE,
  penalized = TRUE,
  alpha = 0.05
)
```

Arguments

x	The residuals or initial data which is a numeric vector or univariate time series
lag	The lag which the statistic will be based on to calculate penalized acf/pacf, default = 1
fitdf	The number of degrees of freedom to be subtracted if x is a series of residuals, default = 0
squared	Take the squared value of residuals to detect nonlinear processes, default=FALSE
log.squared	Take the log of the squared residuals to detect nonlinear processes, default=FALSE
absolute	Take the absolute value of the residuals to detect nonlinear processes, default=FALSE
penalized	If FALSE, perform the unpenalized tests using the sample ACF/PACF estimates, default=TRUE
alpha	The nominal level selected, default = 0.05

Value

A list with class "htest" for penalized/tranditional MahdiMcLeod test results containing statistics, parameters, and p value.

Examples

```
## Not run:
data <- arima.sim(n=100, model=list(ar=0.9))
## Fit AR(1) model
p.fit <- 1
q.fit <- 0
res <- arima(data, order=c(p.fit, 0, q.fit))$resid
out_test1 <- Penalized.MahdiMcLeod.test(x = res, lag=10, fitdf=1, penalized=TRUE)
#'
## Fit ARMA(1,1) model
p.fit <- 1
q.fit <- 1
res <- arima(data, order=c(p.fit, 0, q.fit))$resid
out_test2 <- Penalized.MahdiMcLeod.test(x = res, lag=10, fitdf=1, penalized=TRUE)

## End(Not run)
```

Penalized.LM.test

Description

The Penalized Li-Mak type test

Usage

```
Penalized.LM.test(
  x,
  h.t,
  lag = 1,
  type = c("correlation", "partial"),
  fitdf = 0,
  weighted = TRUE,
  penalized = TRUE,
  alpha = 0.05
)
```

Arguments

x	The residuals or initial data which is a numeric vector or univariate time series
h.t	The sample conditional variances
lag	The lag which the statistic will be based on to calculate penalized acf/pacf, default = 1
type	The type of test based on acf or pacf, default="correlation"
fitdf	The number of degrees of freedom to be subtracted if x is a series of residuals, default = 0
weighted	If FALSE, perform the Li-Mak test, default=TRUE.
penalized	If FALSE, perform the unpenalized tests using the sample ACF/PACF estimates, default=TRUE
alpha	The nominal level selected, default = 0.05

Value

A list with class "htest" for penalized/original Li-Mak tests containing statistics, parameters, and p value.

Examples

```
## Not run:
library(fGarch)
spec <- garchSpec(model=list(ar=0.2,omega=0.2, alpha=c(0.2), beta=0))
x <- garchSim(spec, n=100)
tmp <- garchFit(~arma(1,0)+garch(1,0), data=x, trace=FALSE)
res <- residuals(tmp)
h.t <- attr(tmp, "h.t")
out_test <- Penalized.LM.test(x = res, h.t=h.t, lag=10, type="correlation", fitdf=1, penalized=TRUE)

## End(Not run)
```

Bibliography

Anderson, R. L. (1942). Distribution of the serial correlation coefficient. *The Annals of Mathematical Statistics 13*(1), 1–13.

Andrews, D. W. (1993). Exactly median-unbiased estimation of first order autoregressive/unit root models. *Econometrica: Journal of the Econometric Society*, 139–165.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Arlot, S. and A. Celisse (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys 4*(none), 40 – 79.

Arnau, J. and R. Bono (2001). Autocorrelation and bias in short time series: An alternative estimator. *Quality and Quantity 35*(4), 365–387.

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society 8*(1), 27–41.

Bergmeir, C., R. J. Hyndman, and B. Koo (2018). A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics & Data Analysis 120*, 70–83.

Berk, K. N. (1974, 05). Consistent autoregressive spectral estimates. *Ann. Statist. 2*(3), 489–502.

Bickel, P. J. and Y. R. Gel (2011). Banded regularization of autocovariance matrices in application to parameter estimation and forecasting of time series. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73*(5), 711–728.

Bickel, P. J. and E. Levina (2008). Regularized estimation of large covariance matrices. *Ann. Statist. 36*(1), 199–227.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics 31*(3), 307–327.

Box, G. E. P. and G. M. Jenkins (1976). *Time series analysis: forecasting and control* (Revised ed.). San Francisco, Calif.: Holden-Day. Holden-Day Series in Time Series Analysis.

Box, G. E. P. and D. A. Pierce (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *J. Amer. Statist. Assoc. 65*, 1509–1526.

Brockwell, P. J. and R. A. Davis (1991). *Time series: theory and methods* (Second ed.). Springer Series in Statistics. New York: Springer-Verlag.

Brockwell, P. J. and R. A. Davis (2002). *Introduction to time series and forecasting* (Second ed.). Springer Texts in Statistics. New York: Springer-Verlag. With 1 CD-ROM (Windows).

Burman, P., E. Chow, and D. Nolan (1994). A cross-validatory method for dependent data. *Biometrika 81*(2), 351–358.

Burman, P. and D. Nolan (1992). Data-dependent estimation of prediction functions. *Journal of Time Series Analysis 13*(3), 189–207.

Carnero, M. A., D. Peña, and E. Ruiz (2001). Outliers and conditional autoregressive heteroscedasticity in time series. *Estadística 53*, 143–213.

Chang, C. C. and D. N. Politis (2016). Robust autocorrelation estimation. *Journal of Computational and Graphical Statistics 25*(1), 144–166.

Chu, C.-K. and J. S. Marron (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics 19*(4), 1906–1918.

DeCarlo, L. T. and W. W. Tryon (1993). Estimating and testing autocorrelation with small samples: A comparison of the c-statistic to a modified estimator. *Behaviour research and therapy 31*(8), 781–788.

Durbin, J. (1960). The fitting of time-series models. *Revue de l'Institut International de Statistique*, 233–244.

Dürre, A., R. Fried, and T. Liboschik (2015). Robust estimation of (partial) autocorrelation. *Wiley Interdisciplinary Reviews: Computational Statistics 7*(3), 205–222.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica 50*(4), 987–1007.

Fan, J. (2004, 03). Generalised likelihood ratio tests for spectral density. *Biometrika 91*, 195–209.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association 96*(456), 1348–1360.

Fan, J. and Q. Yao (2003). *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer.

Finkenstädt, B. (2012). *Nonlinear dynamics in economics: a theoretical and statistical approach to agricultural markets*, Volume 426. Springer Science & Business Media.

Fisher, T. J. and C. M. Gallagher (2012). New weighted portmanteau statistics for time series goodness of fit testing. *J. Amer. Statist. Assoc. 107*(498), 777–787.

Fisher, T. J. and M. W. Robbins (2018). An improved measure for lack of fit in time series models. *Statistica Sinica 28*(3), 1285–1305.

Fuller, W. A. (1976). *Introduction to statistical time series*. New York: Wiley.

Gallagher, C. M. and T. J. Fisher (2015). On weighted portmanteau tests for time-series goodness-of-fit. *Journal of Time Series Analysis 36*(1), 67–83.

Györfi, L., W. Härdle, P. Sarda, and P. Vieu (1989). How to select the smoothing parameter? In *Nonparametric Curve Estimation from Time Series*, pp. 105–133. Springer.

Hart, J. D. and T. E. Wehrly (1986). Kernel regression estimation using repeated measurements data. *Journal of the American Statistical Association 81*(396), 1080–1088.

Higgins, M. L. and A. K. Bera (1992). A class of nonlinear arch models. *International Economic Review 33*(1), 137–58.

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

Huitema, B. E. and J. W. McKean (1991). Autocorrelation estimation and inference with small samples. *Psychological Bulletin 110*(2), 291.

Huitema, B. E. and J. W. McKean (1994). Two reduced-bias autocorrelation estimators: rf1 and rf2. *Perceptual and Motor Skills 78*(1), 323–330.

Hyndman, R. J. (2014). Measuring forecast accuracy. *Business forecasting: Practical problems and solutions*, 177–183.

Keenan, D. M. (1985). A tukey nonadditivity-type test for time series nonlinearity. *Biometrika 72*, 39–44.

Kendall, M. and J. Ord (1990). Time series, i edward arnold. *Sevenoaks, Kent, Great Britain*, 51–52.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika 41*(3-4), 403–404.

Li, G. and W. K. Li (2005). Diagnostic checking for time series models with conditional heteroscedasticity estimated by the least absolute deviation approach. *Biometrika 92*(3), 691–701.

Li, W. K. and T. K. Mak (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *J. Time Ser. Anal. 15*(6), 627–636.

Li, W. K. and A. I. McLeod (1981). Distribution of the residual autocorrelations in multivariate ARMA time series models. *J. Roy. Statist. Soc. Ser. B 43*(2), 231–239.

Liao, L., C. Park, J. Hannig, and K.-H. Kang (2016). Autocovariance function estimation via penalized regression. *Journal of Computational and Graphical Statistics 25*(4), 1041–1056.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

MacKinnon, J. G. and A. A. Smith Jr (1998). Approximate bias correction in econometrics. *Journal of Econometrics 85*(2), 205–230.

Mahdi, E. (2016). Portmanteau test statistics for seasonal serial correlation in time series models. *SpringerPlus 5*(1), 1485.

Mahdi, E. and A. I. McLeod (2012). Improved multivariate portmanteau test. *J. Time Series Anal. 33*(2), 211–222.

Marriot, F. H. C. and J. A. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika 41*(3-4), 390–402.

Marriott, F. and J. Pope (1954). Bias in the estimation of autocorrelations. *Biometrika 41*(3/4), 390–402.

McLeod, A. I. and W. K. Li (1983a). Diagnostic checking arma time series models using squared-residual autocorrelations. *Journal of time series analysis 4*(4), 269–273.

McLeod, A. I. and W. K. Li (1983b). Diagnostic checking ARMA time series models using squared-residual autocorrelations. *J. Time Ser. Anal. 4*(4), 269–273.

McMurry, T. L. and D. N. Politis (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis 31*(6), 471–482.

McMurry, T. L. and D. N. Politis (2015). High-dimensional autocovariance matrices and optimal linear prediction. *Electronic Journal of Statistics 9*(1), 753–788.

Monti, A. C. (1994). A proposal for a residual autocorrelation test in linear models. *Biometrika 81*(4), 776–780.

Moran, P. A. (1948). Some theorems on time series: Ii the significance of the serial correlation coefficient. *Biometrika 35*(3/4), 255–260.

Müller, U. K. (2014). Hac corrections for strongly autocorrelated time series. *Journal of Business & Economic Statistics 32*(3), 311–322.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*(3), 703–708.

Opsomer, J., Y. Wang, and Y. Yang (2001). Nonparametric regression with correlated errors. *Statistical Science*, 134–153.

Orcutt, G. H. and J. Irwin (1948). A study of the autoregressive nature of the time series used for tinbergen's model of the economic system of the united states, 1919-1932. *Journal of the Royal Statistical Society. Series B (Methodological) 10*(1), 1–53.

Peña, D. and J. Rodríguez (2002). A powerful portmanteau test of lack of fit for time series. *J. Amer. Statist. Assoc. 97*(458), 601–610.

Peña, D. and J. Rodríguez (2006). The log of the determinant of the autocorrelation matrix for testing goodness of fit in time series. *J. Statist. Plann. Inference 136*(8), 2706–2718.

Pérez, A. and E. Ruiz (2003). Properties of the sample autocorrelations of nonlinear transformations in long-memory stochastic volatility models. *Journal of Financial Econometrics 1*(3), 420–444.

Proietti, T. and A. Giovannelli (2018). A durbin–levinson regularized estimator of high-dimensional autocovariance matrices. *Biometrika 105*(4), 783–795.

Psaradakis, Z. and M. Vávra (2019). Portmanteau tests for linearity of stationary time series. *Econometric Reviews 38*(2), 248–262.

Racine, J. (2000). Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of econometrics 99*(1), 39–61.

Reschenhofer, E. (2019). Heteroscedasticity-robust estimation of autocorrelation. *Communications in Statistics-Simulation and Computation 48*(4), 1251–1263.

Rodríguez, J. and E. Ruiz (2005). A powerful test for conditional heteroscedasticity for financial time series with highly persistent volatilities. *Statist. Sinica 15*(2), 505–523.

Solanas, A., R. Manolov, and V. Sierra (2010). Lag-one autocorrelation in short series: Estimation and hypotheses testing. *Psicológica 31*(2), 357–381.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the royal statistical society: Series B (Methodological) 36*(2), 111–133.

Taylor, S. J. (1986). *Modelling Financial Time Series*. New York: Wiley.

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

Tol, R. S. (1996). Autoregressive conditional heteroscedasticity in daily temperature measurements. *Environmetrics 7*(1), 67–75.

Tse, Y. K. (2002). Residual-based diagnostics for conditional heteroscedasticity models. *Econom. J. 5*(2), 358–373.

Tsui, A. K. (2004). Diagnostics for conditional heteroscedasticity models: some simulation results. *Math. Comput. Simulation 64*(1), 113–119. First special issue: selected papers of the MSSANZ/IMACS 14th Biennial Conference on Modelling and Simulation (Canberra, 2001).

Tuan, D. P. (1992). Approximate distribution of parameter estimators for first-order autoregressive models. *Journal of Time Series Analysis 13*(2), 147–170.

Ulrych, T. J. and R. W. Clayton (1976). Time series modelling and maximum entropy. *Physics of the Earth and Planetary Interiors 12*(2-3), 188–200.

White, J. S. (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. *Biometrika 48*(1-2), 85–94.

Wooldridge, J. M. (1991). On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *J. Econometrics 47*(1), 5–46.

Wu, W. B. and M. Pourahmadi (2003). Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika 90*(4), 831–844.

Wu, W. B. and M. Pourahmadi (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 1755–1768.

Young, L. (1941). On randomness in ordered sequences. *The Annals of Mathematical Statistics 12*(3), 293–300.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association 101*(476), 1418–1429.