

# Autocovariance Function Estimation via Penalized Regression

LINA LIAO, CHEOLWOO PARK

Department of Statistics, University of Georgia, Athens, GA 30602, USA

JAN HANNIG

Department of Statistics and Operations Research,

University of North Carolina, Chapel Hill, NC 27599, USA

KEE-HOON KANG

Department of Statistics, Hankuk University of Foreign Studies, Yongin, 449-791, Korea

## Abstract

The work revisits the autocovariance function estimation, a fundamental problem in statistical inference for time series. We convert the function estimation problem into constrained penalized regression with a generalized penalty that provides us with flexible and accurate estimation, and study the asymptotic properties of the proposed estimator. In case of a nonzero mean time series, we apply a penalized regression technique to a differenced time series, which does not require a separate detrending procedure. In penalized regression, selection of tuning parameters is critical and we propose four different data-driven criteria to determine them. A simulation study shows effectiveness of the tuning parameter selection and that the proposed approach is superior to three existing methods. We also briefly discuss the extension of the proposed approach to interval-valued time series.

*Some key words:* Autocovariance function; Differenced time series; Regularization; Time series; Tuning parameter selection.

# 1 Introduction

Let  $\{Y_t, t \in T\}$  be a stochastic process such that  $Var(Y_t) < \infty$ , for all  $t \in T$ . The autocovariance function of  $\{Y_t\}$  is given as

$$\gamma(s, t) = Cov(Y_s, Y_t)$$

for all  $s, t \in T$ . In this work, we consider a regularly sampled time series  $\{(i, Y_i), i = 1, \dots, n\}$ . Its model can be written as

$$Y_i = g(i) + \epsilon_i, \quad i = 1, \dots, n, \tag{1.1}$$

where  $g$  is a smooth deterministic function and the error is assumed to be a weakly stationary process with  $E(\epsilon_i) = 0$ ,  $Var(\epsilon_i) = \sigma^2$  and  $Cov(\epsilon_i, \epsilon_j) = \gamma(|i - j|)$  for all  $i, j = 1, \dots, n$ .

The estimation of the autocovariance function  $\gamma$  (or autocorrelation function) is crucial to determine the degree of serial correlation in a time series. It also plays a key role for statistical inference in time series analysis. For a mean zero time series  $\{Y_i\}$  with  $g = 0$  in (1.1), the autocovariance function can be estimated by its simple sample estimator:

$$\bar{\gamma}(k) = \frac{1}{n} \sum_{i=1}^{n-k} Y_i Y_{i+k}, \quad k = 0, 1, \dots, n - 1. \tag{1.2}$$

An extensive literature exists to improve the sample autocovariance estimator in time series analysis. Hall et al. (1994) estimate the covariance function of a stationary process, which ensures the positive semi-definiteness of the estimator. Hyndman and Wand (1997) propose a bias adjustment approach, and Fan and Kreutzberger (1998) estimate the spectral density function using a local polynomial regression. Huang et al. (2011) also estimate the spectral density function based on a regularized inverse technique. Wu and Pourahmadi (2009) propose a banded, not necessarily positive definite, covariance matrix estimator and calculate its explicit rates of convergence under a short range dependence.

In this work, we propose to estimate an autocovariance function by converting the function estimation problem into constrained penalized regression with a generalized penalty. Regularization techniques enable us to obtain more stable and interpretable model estimates because structural information within the data can be incorporated into the model estimation procedure to significantly reduce the actual complexity involved in the estimation procedure. We show that the proposed estimator shares good asymptotic properties with the sample autocovariance estimator in (1.2). In case of a nonzero mean time series, we

apply a penalized regression technique to the differenced time series, which does not require a separate detrending procedure. In penalized regression, there are tuning parameters that should be determined from the observed time series. We propose four different data-driven criteria to select these parameters using generalized cross-validation and L-curve approaches.

In Section 2, we briefly review three existing autocovariance estimators whose finite sample performances are compared in Section 4. In Section 3, we propose our estimators for both zero mean and nonzero mean time series. For a special case of the proposed method, we obtain the closed form of the estimator and study its asymptotic normality and consistency. We also describe generalized cross-validation and L-curve approaches to select tuning parameters in constrained penalized regression. Section 4 presents the numerical comparison of the proposed method with the three existing methods introduced in Section 2. We also discuss the nonnegative definiteness of the proposed estimator. Section 5 describes possible extension of the proposed work to interval-valued time series. In Section 6, the technical proofs of theoretical properties of the proposed estimator are presented.

## 2 Literature Review

In this section, we introduce three existing estimators of  $\gamma$ , which assume zero mean, i.e.  $g = 0$  in (1.1). We compare the proposed method with these three methods in Section 4.

Huang et al. (2011) develop the spectral density estimator through a regularized inverse problem. They consider a real-valued second-order stationary process  $Y = \{Y_t, t \in T\}$  with zero mean and covariance function  $R$ :

$$R(t) = \int e^{it\omega} f(\omega) d\omega = 2 \int_0^\infty \cos(t\omega) f(\omega) d\omega.$$

Then, one can obtain the spectral density  $f$

$$f(\omega) = (2\pi)^{-1} \int e^{-it\omega} R(t) dt. \tag{2.1}$$

Unlike traditional approaches, Huang et al. (2011) do not attempt to estimate the spectral density  $f$  from (2.1) or from the periodogram. Instead, they propose to estimate  $f$  from an estimate of  $R$  by solving a regularized inverse problem. Since  $E(Y) = 0$ ,  $E(Y_{t_i} Y_{t_j}) = R(t_i - t_j)$ . In other words,  $Y_{t_i} Y_{t_j}$  is an unbiased estimator of

$$R(t_i - t_j) = 2 \int_0^\infty \cos((t_i - t_j)\omega) f(\omega) d\omega.$$

Then, by focusing on  $[0, \nu]$  to avoid the aliasing effect, they propose an estimator of  $f$  as

$$\hat{f}_\lambda(\omega) = \begin{cases} \frac{1}{\nu} \hat{f}^*_\lambda(\frac{\omega}{\nu}), & \omega \in [0, \nu], \\ 0, & \text{otherwise,} \end{cases}$$

where  $\hat{f}^*_\lambda$  minimizes the following objective function:

$$\sum_{1 \leq i, j \leq n} [Y_{t_i} Y_{t_j} - \int_0^1 \cos(\nu(t_i - t_j)u) f^*(u) du]^2 + \lambda \int_0^1 [f^{*(m)}(u)]^2 du,$$

where  $\lambda > 0$  is a tuning parameter. They consider four different ways of choosing  $\lambda$  using generalized cross-validation.

Wu and Pourahmadi (2009) apply an idea of banding a covariance matrix and propose the following estimator:

$$\hat{\Sigma}_{n,l} = (\bar{\gamma}(|i - j|) \mathbf{1}_{|i-j| \leq l})_{1 \leq i, j \leq n}$$

where  $\bar{\gamma}(k), k = 0, 1, \dots, n - 1$ , are the sample autocovariance estimator in (1.2) and  $l \geq 0$  is an integer. They show that the banded covariance matrix estimator converges to the true covariance matrix under a short-range dependence assumption for a wide class of nonlinear processes. However, their estimator does not guarantee the positive definiteness.

Fan and Kreutzberger (1998) utilize a local smoothing technique for the spectral density estimation based on the Whittle likelihood estimator of a spectral density (Whittle, 1962). Let  $\omega_l = 2\pi l/n$  ( $l = 0, 1, \dots, N = [(n - 1)/2]$ ). For each given  $\omega$ , they form the weighted log-likelihood as follows:

$$L(a, b) = \sum_{l=1}^N [-\exp\{Y_l - a - b(\omega_l - \omega)\} + Y_l - a - b(\omega_l - \omega)] K_h(\omega_l - \omega), \quad (2.2)$$

where  $K_h(\cdot) = K(\cdot/h)/h$ . Then, the proposed local likelihood estimator for the logarithm of the spectral density  $m(\omega) = \log f(\omega)$  is given as  $\hat{m}(\omega) = \hat{a}$  where  $\hat{a}$  and  $\hat{b}$  are the maximizers of (2.2).

### 3 Regularized Autocovariance Function Estimation

In this section, we propose a new regularized estimator of an autocovariance function by assuming that we observe a regularly spaced time series in (1.1). Unlike the approach in

Huang et al. (2011), we apply a regularized technique to the time domain. In Section 3.1, we develop an estimator for a zero mean time series, and investigate its theoretical properties for a special case. In Section 3.2, we propose a difference-based estimator for a nonzero mean time series, which does not require a separate detrending procedure. In Section 3.3, we suggest four different criteria using generalized cross-validation and L-curve techniques to select tuning parameters in penalized regression.

### 3.1 Time Series with Zero Mean

For a time series with zero mean, i.e.  $g = 0$  in (1.1),

$$\text{Cov}(Y_i, Y_j) = E(Y_i Y_j) = \gamma(|i - j|),$$

for all  $1 \leq i, j \leq n$ . Then, using this relationship, one can set up a regression model:

$$Y_i Y_j = \gamma(|i - j|) + \delta_{ij},$$

where  $E(\delta_{ij}) = 0$ . Noting that estimating  $\gamma(|i - j|)$  for  $1 \leq i, j \leq n$  by the least squares method does not lead to a stable solution, we apply a regularization technique with constraints to the estimation problem. First, since  $\gamma(0) \geq |\gamma(k)|$  for  $k = 1, \dots, n - 1$ , we consider only such solutions. Second, we introduce the generalized penalty  $\lambda \sum_{k=1}^{n-1} k^\alpha |\gamma(k)|^q$  where  $\lambda, \alpha, q > 0$  by assuming that serial correlation in a time series quickly decreases as a time lag increases. Hence, the proposed method works best for a weakly correlated time series. The weight  $k^\alpha$  reflects the fact that  $\gamma(k)$  should decay as the lag  $k$  increases. Here,  $\alpha$  represents the decay degree of the autocovariance function, and  $q$  controls the norm of the regression coefficients. Note that the penalty does not apply to  $\gamma(0)$ . When  $q = 1$  and  $q = 2$ , the estimation corresponds to LASSO (Tibshirani, 1996) and ridge penalties (Hoerl and Kennard, 1970), respectively, when  $\alpha = 0$ . The tuning parameter  $\lambda$  trades off between the loss function and the penalty function. As a result, we obtain the estimator  $\hat{\boldsymbol{\gamma}}$  of  $\boldsymbol{\gamma} = (\gamma(0), \gamma(1), \dots, \gamma(n - 1))^T$  by minimizing

$$f_n(\boldsymbol{\gamma}; \lambda, \alpha, q) = \sum_{i \leq j} \left( Y_i Y_j - \gamma(|i - j|) \right)^2 + \lambda \sum_{k=1}^{n-1} k^\alpha |\gamma(k)|^q \quad (3.1)$$

on  $\mathcal{C} = \{\boldsymbol{\gamma} : \gamma(0) \geq |\gamma(k)|, k = 1, \dots, n - 1\}$ .

For the case of  $\alpha = q = 2$ , one can rewrite  $f_n(\boldsymbol{\gamma}; \lambda)$  in (3.1) in the following matrix form:

$$f_n(\boldsymbol{\gamma}; \lambda) = (\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}) + \lambda (\mathbf{A}\boldsymbol{\gamma})^T (\mathbf{A}\boldsymbol{\gamma}).$$

Here,  $\mathbf{y} = (Y_1^2, Y_2^2, \dots, Y_n^2, Y_1Y_2, Y_2Y_3, \dots, Y_{n-1}Y_n, \dots, Y_1Y_n)^\top = (\mathbf{z}_0^\top, \mathbf{z}_1^\top, \dots, \mathbf{z}_{n-1}^\top)^\top$  is an  $n(n+1)/2 \times 1$  vector,  $\mathbf{z}_k$  is an  $(n-k) \times 1$  vector,  $k = 0, 1, \dots, n-1$ ;  $\mathbf{X} = \text{diag}(\mathbf{1}_n, \mathbf{1}_{n-1}, \dots, 1)$  is an  $n(n+1)/2 \times n$  matrix,  $\mathbf{1}_k$  is an  $k \times 1$  vector with all entries 1,  $k = 1, \dots, n$ ; and  $\mathbf{A} = \text{diag}(0, 1, \dots, n-1)$  is a  $n \times n$  matrix.

One can obtain the explicit solution for fixed  $\alpha = q = 2$ ; we solve the equations of derivatives of  $f_n(\boldsymbol{\gamma}; \lambda)$  with respect to  $\gamma(k)$ ,  $\frac{\partial f_n(\boldsymbol{\gamma}; \lambda)}{\partial \gamma(k)} = 0$ ,  $k = 0, 1, \dots, n-1$ . This leads to

$$\hat{\gamma}(0) = \frac{1}{n} \mathbf{1}_n^\top \mathbf{z}_0, \quad \hat{\gamma}(k) = \frac{1}{(n-k) + \lambda k^2} \mathbf{1}_{n-k}^\top \mathbf{z}_k, \quad \text{for } k = 1, \dots, n-1. \quad (3.2)$$

We show the derivation of the solutions in (3.2) and that they satisfy the restriction  $\mathcal{C} = \{\gamma : \gamma(0) \geq |\gamma(k)|, k = 1, \dots, n-1\}$  in Section 6. Note that when  $\lambda = 0$ , the solution becomes

$$\tilde{\gamma}(k) = \frac{1}{n-k} \sum_{i=1}^{n-k} Y_i Y_{i+k}, \quad k = 0, 1, \dots, n-1,$$

which is an alternative sample autocovariance estimator.

In what follows, we present theoretical properties of the estimator in (3.2). To state the theorems, we adopt the notations from Wu (2005). We assume that  $Y_i$  is of form  $Y_i = \tilde{g}(\dots, \tilde{\epsilon}_{i-1}, \tilde{\epsilon}_i)$ , where  $\tilde{\epsilon}_i$  are i.i.d. innovations or shocks that drive the system,  $\mathcal{F}_n = (\dots, \tilde{\epsilon}_{n-1}, \tilde{\epsilon}_n)$  is the input,  $\tilde{g}(\cdot)$  is a real-valued function and  $Y_n = \tilde{g}(\mathcal{F}_n)$  is the output. Let  $(\tilde{\epsilon}'_i)_{i \in \mathbb{Z}}$  be an i.i.d. copy of  $(\tilde{\epsilon}_i)_{i \in \mathbb{Z}}$ . Then, the physical dependence measure is defined as  $\delta_p(n) = \|Y_n - Y'_n\|_p$ , where  $Y'_n = g(\mathcal{F}'_n)$ ,  $\mathcal{F}'_n = (\mathcal{F}_{-1}, \tilde{\epsilon}'_0, \tilde{\epsilon}_1, \dots, \tilde{\epsilon}_n)$ . If  $\|Z\|_p \equiv [E(|Z|^p)]^{1/p} < \infty$  for  $p > 0$ , then  $Z$  is said to belong to  $\mathcal{L}^p$ , i.e.,  $Z \in \mathcal{L}^p$ . We also assume that  $E(Y_i) = 0$  and

$$\Delta_p \equiv \sum_{i=0}^{\infty} \delta_p(i) < \infty. \quad (3.3)$$

The first theorem states the asymptotic normality of the proposed estimator. For fixed  $k$ ,  $0 \leq k \leq n-1$ , let  $\boldsymbol{\eta}_i = (Y_i, Y_{i-1}, \dots, Y_{i-k})^\top$ ,  $\mathbf{A}_k = \text{diag}(1, \frac{n}{(n-1)+\lambda}, \dots, \frac{n}{(n-k)+\lambda k^2})$ . Also, let  $\hat{\boldsymbol{\Gamma}}_k$  be the proposed estimator of  $\boldsymbol{\Gamma}_k = (\gamma(0), \gamma(1), \dots, \gamma(k))^\top$ .

**Theorem 1.** *Assume  $Y_i \in \mathcal{L}^4$  and (3.3) holds with  $p = 4$ . Then,*

$$\sqrt{n}(\hat{\boldsymbol{\Gamma}}_k - \boldsymbol{\Gamma}_k) = \frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i \mathbf{A}_k \boldsymbol{\eta}_i - \boldsymbol{\Gamma}_k) \rightarrow N(\mathbf{0}, E(\mathbf{D}_0 \mathbf{D}_0^\top)), \quad k = 0, 1, \dots, n-1. \quad (3.4)$$

where  $\mathbf{D}_0 = \sum_{i=0}^{\infty} \mathcal{P}_0(Y_i \mathbf{A}_k \boldsymbol{\eta}_i) \in \mathcal{L}^2$  and  $\mathcal{P}_0$  is the projection operator defined by

$$\mathcal{P}_i(\cdot) = E(\cdot | \mathcal{F}_i) - E(\cdot | \mathcal{F}_{i-1}), \quad i \in \mathbb{Z}.$$

The next theorem states the upper bound of the difference between the proposed estimator and the (scaled) true autocovariance function at each lag.

**Theorem 2.** *Assume  $Y_i \in \mathcal{L}^p$ ,  $2 < p \leq 4$ , and (3.3) holds with  $2 < p \leq 4$ . Then, we have*

$$\left\| \hat{\gamma}(k) - \left( 1 - \frac{\lambda k^2}{(n-k) + \lambda k^2} \right) \gamma(k) \right\|_{p/2} \leq \frac{4 \|Y_1\|_p \Delta_p}{(p-2)n^{1-2/p}}, \quad k = 0, 1, \dots, n-1.$$

The proofs for the two theorems are delayed to Section 6.

We note that the proposed estimator does not warrant the nonnegative definiteness. As a practical solution, we suggest to replace negative eigenvalues of the estimate with zero values and revise the estimate. More specifically, applying singular value decomposition to the original estimate, we obtain its eigenvalues. Then, we replace the negative ones with zero and reconstruct the covariance estimate, which would result in nonnegative definiteness. A simulation study in Section 4.2 shows that there are minimal differences between the original and revised estimates in terms of estimation.

### 3.2 Time Series with Nonzero Mean

For a time series with nonzero mean, we note that the methods introduced in Section 2 would require a separate estimation of the trend function  $g$  in (1.1), which is a nontrivial task and might introduce a bias in the estimation of the autocovariance function (Park et al., 2009).

We propose the autocovariance function estimation that utilizes constrained penalized regression for a differenced time series. We use the first difference in this paper, but it can be generalized to a higher order differencing. Let  $e_i$  be the first differenced time series, i.e.,  $\mathbf{e} = \mathbf{D}\mathbf{Y}^o$  where  $\mathbf{Y}^o = (Y_1, \dots, Y_n)^\top$  and  $\mathbf{D} = (a_{i,k})$  is the first difference matrix, i.e.,  $a_{i,i} = -1$  and  $a_{i,i+1} = 1$  for  $i = 1, \dots, n-1$ , and the other elements are zero. Then, one can derive the following relationship (Park et al., 2009):

$$\text{Cov}(e_i, e_j) = \sum_{k=1}^n a_{i,k} a_{j,k} \gamma(0) + \sum_{k=1}^{n-1} (a_{i,k} a_{j,k+1} + a_{i,k+1} a_{j,k}) \gamma(1) + \dots + (a_{i,1} a_{j,n} + a_{i,n} a_{j,1}) \gamma(n-1),$$

for all  $i$  and  $j$ . One can set up the following regression model:

$$e_i e_j = \sum_{k=1}^n a_{i,k} a_{j,k} \gamma(0) + \sum_{k=1}^{n-1} (a_{i,k} a_{j,k+1} + a_{i,k+1} a_{j,k}) \gamma(1) + \dots + (a_{i,1} a_{j,n} + a_{i,n} a_{j,1}) \gamma(n-1) + \delta_{ij},$$

if a regression function is smooth such that  $E(\delta_{ij}) \approx 0$ . As a result, we have the following objective function

$$f_n^*(\boldsymbol{\gamma}; \lambda, \alpha, q) = \sum_{i \leq j} \left( e_i e_j - \sum_{k=1}^n a_{i,k} a_{j,k} \gamma(0) - \sum_{k=1}^{n-1} (a_{i,k} a_{j,k+1} + a_{i,k+1} a_{j,k}) \gamma(1) - \cdots - (a_{i,1} a_{j,n} + a_{i,n} a_{j,1}) \gamma(n-1) \right)^2 + \lambda \sum_{i=1}^{n-1} i^\alpha |\gamma(i)|^q, \quad (3.5)$$

for a differenced time series. We estimate  $\boldsymbol{\gamma}$  using

$$\arg \min_{\boldsymbol{\gamma} \in \mathcal{C}} f_n^*(\boldsymbol{\gamma}; \lambda, \alpha, q)$$

where  $\mathcal{C} = \{\boldsymbol{\gamma} : \gamma(0) \geq |\gamma(k)|, k = 1, \dots, n-1\}$ .

Park et al. (2009) investigate a special case of  $\alpha = 2$ ,  $q = 2$ , and  $\lambda = 1$  for a nonzero mean time series. They show that the method provides more stable estimates compared to the sample autocovariance estimator. In this work, we treat  $\lambda$ ,  $\alpha$  and  $q$  as tuning parameters and search for optimal values which produce the smallest estimation error. After exhaustive search we learn that  $\alpha$  values do not make critical differences in the performance of the proposed estimator. Therefore, we fix  $\alpha = 2$  in our simulation study. In the following subsection, we suggest four criteria to choose  $\lambda$  and  $q$ .

**Remark.** For an unevenly sampled time series, a differenced time series  $\{e_i\}$  cannot be defined. One might use a difference operator  $\psi$  such as wavelets or the first derivative of Gaussian function. For example, let  $\psi_{j,l}(x) = 2^{j/2} \psi(2^j x - l)$  be a dilated and translated wavelet function. The wavelet coefficients are defined as

$$d_{j,l} = \int \psi_{j,l}(u) Y(u) du \approx \frac{1}{n} \sum_{i=1}^n \psi_{j,l}(t_i) Y(t_i),$$

and we note that  $E(d_{j,l}) = 0$ . Then, for a given  $l_1$  and  $l_2$ ,

$$Cov(d_{j,l_1}, d_{j,l_2}) \approx \frac{1}{n^2} \sum_{i=1}^n \sum_{m=1}^n \psi_{j,l_1}(t_i) \psi_{j,l_2}(t_m) \gamma(|t_i - t_m|).$$

Using this relationship, one can construct a similar constrained penalized regression setting based on the wavelet coefficients to estimate the autocovariance function  $\gamma$ :

$$\arg \min_{\boldsymbol{\gamma}} \sum_{l_1} \sum_{l_2} \left( d_{j,l_1} d_{j,l_2} - \frac{1}{n^2} \sum_{i=1}^n \sum_{m=1}^n \psi_{j,l_1}(t_i) \psi_{j,l_2}(t_m) \gamma(|t_i - t_m|) \right)^2 + p_{\lambda, \alpha, q}(\boldsymbol{\gamma})$$

under the condition that  $\gamma(0) > \gamma(t)$  for any  $t > 0$ . Here,  $p_{\lambda, \alpha, q}(\boldsymbol{\gamma})$  represents a penalty function on  $\boldsymbol{\gamma}$ . We suggest this study as our future work.



### 3.3 Selection of Tuning Parameters

We first fix  $q = 2$  and illustrate how to choose  $\lambda$  using generalized cross-validation (GCV), L-curve, and BIC, and then discuss the choice of  $q$  later.

For a time series with zero mean, the GCV function is defined as

$$GCV(\lambda) = \frac{\frac{1}{n} \|(\mathbf{I} - \mathbf{H}_\lambda)\mathbf{y}\|^2}{\left[\frac{1}{n} \text{tr}(\mathbf{I} - \mathbf{H}_\lambda)\right]^2}$$

where  $\mathbf{H}_\lambda$  is the hat matrix, defined by

$$\mathbf{H}_\lambda = \mathbf{X}(\mathbf{X}^\top \mathbf{X} + \lambda \mathbf{A}^\top \mathbf{A})^{-1} \mathbf{X}^\top.$$

One can select an optimal  $\lambda$  by minimizing  $GCV(\lambda)$ . As pointed out in Huang et al. (2011), however, when  $\{Y_i\}$  are highly correlated, GCV might perform poorly. Therefore, we also consider the modified version of GCV in Huang et al. (2011), defined by

$$V(\lambda) = \frac{1}{n} \|(\mathbf{I} - \mathbf{H}_\lambda)\mathbf{y}\|^2 - \frac{1}{n} \text{tr}(\mathbf{S}) + \frac{2}{n} \text{tr}(\mathbf{S}\mathbf{H}_\lambda)$$

where  $\mathbf{S}$  is an unbiased estimator of the covariance matrix of  $\mathbf{y}$ . One would instead select  $\lambda$  as the minimizer of  $V(\lambda)$ .

In what follows we briefly introduce the L-curve approach. Let  $\boldsymbol{\gamma}_\lambda = \arg \min_{\boldsymbol{\gamma} \in \mathcal{C}} f_n(\boldsymbol{\gamma}; \lambda)$ . L-curve is a plot that displays the relationship between  $\|\mathbf{X}\boldsymbol{\gamma}_\lambda - \mathbf{y}\|$  and  $\|\mathbf{A}\boldsymbol{\gamma}_\lambda\|$ , i.e. between loss and penalty functions. Hansen (2000) argues that a typical L-curve plot would show both horizontal and vertical parts of the curve where the solution is contributed by the penalty and loss function, respectively. Hence, in order to trade off these two parts, an optimal position in a plot would be a corner where the curve levels off. After the standardization transformation (denoted as  $\bar{\cdot}$  below) introduced in Hansen (1987), L-curve aims to find the best tradeoff between  $\|\bar{\mathbf{X}}\bar{\boldsymbol{\gamma}}_\lambda - \bar{\mathbf{y}}\|$  and  $\|\bar{\boldsymbol{\gamma}}_\lambda\|$ . The criterion we use is the curvature of the L-curve. Let  $\hat{\eta} = \log \|\bar{\boldsymbol{\gamma}}_\lambda\|^2$  and  $\hat{\rho} = \log \|\bar{\mathbf{X}}\bar{\boldsymbol{\gamma}}_\lambda - \bar{\mathbf{y}}\|^2$ . The curvature  $\kappa$  of the L-curve is defined as

$$\kappa = 2 \frac{\hat{\rho}' \hat{\eta}'' - \hat{\rho}'' \hat{\eta}'}{((\hat{\rho}')^2 + (\hat{\eta}')^2)^{3/2}}$$

where  $\hat{\rho}', \hat{\eta}', \hat{\rho}''$  and  $\hat{\eta}''$  are the first and second derivatives of  $\hat{\rho}$  and  $\hat{\eta}$  with respect to  $\lambda$ . Then, we choose  $\lambda$  that maximizes  $\kappa$ , which corresponds to the corner position in a L-curve plot.

Besides GCV, we also apply BIC, which is a popular tuning parameter selection approach. We find the optimal  $\lambda$  using the following criterion (Zou et al., 2007):

$$\text{BIC} = \log \left( \frac{1}{n} \|(\mathbf{I} - \mathbf{H}_\lambda)\mathbf{y}\|^2 \right) + \hat{\text{df}} \log n/n$$

where  $\hat{d}f = \text{tr}(\mathbf{H}_\lambda)$ .

For a nonzero mean times series, the same procedures, (modified) GCV, L-curve, and BIC, can be applied to the differenced times series.

One would expect that serial correlation should disappear after  $m$  lags for a weakly correlated time series, which would result in  $\gamma(k) = 0$  for  $k > m$ . In this sense, sparsity in a regression model might be an important property that a desirable autocovariance estimator should possess. Therefore, we also implement constrained regularized regression with  $0 < q \leq 1$  in both (3.1) and (3.5) to achieve sparsity. We treat  $q$  in the range between 0 and 1 as another tuning parameter along with  $\lambda$ . In this case, we notice that both (modified) GCV and BIC do not provide a reasonable solution due to singularity in the hat matrix,  $\mathbf{H}_\lambda$ , and thus we choose the optimal combination of  $(q, \lambda)$  that minimizes L-curve.

## 4 A Simulation Study

In this section, we conduct a simulation study using various time series models with length  $n = 100$ . Four different simulated examples are considered:

- The zero mean  $MA(1)$  model:  $Y_i = \epsilon_i + \theta\epsilon_{i-1}$ . We take  $\theta = 0.1, 0.5, 0.9$ .
- The zero mean  $MA(2)$  model:  $Y_t = \epsilon_i + \theta_1\epsilon_{i-1} + \theta_2\epsilon_{i-2}$ . We take  $\theta_1 = 0.9$  and  $\theta_2 = 0.8$ .
- The zero mean  $AR(1)$  model:  $Y_i = \phi Y_{i-1} + \epsilon_i$ . We take  $\phi = 0.1, 0.5, 0.9$ .
- The zero mean  $AR(2)$  model:  $Y_i = \phi_1 Y_{i-1} + \phi_2 Y_{i-2} + \epsilon_i$ . We take  $\phi_1 = -0.9$  and  $\phi_2 = 0.2$ .

Throughout this section, we assume  $\{\epsilon_i\}$  follows the standard normal distribution. We also consider nonzero mean MA and AR models by adding the signal  $g(i) = \sin(3 \times 2\pi i/n)$ . We generate both training and testing data sets with 100 replications, and measure the mean squared error (MSE) between the estimated and true  $\gamma$  for each simulation setting, for each method, and for both training and testing data sets. For the nonzero mean cases, except the proposed method, local linear regression is applied to remove the mean function  $g$  (Fan and Kreutzberger, 1998). For the proposed method, the original time series  $\{Y_i\}$  are used for the zero mean cases and the differenced time series  $\{e_i\}$  are used for the nonzero mean cases.

We compare the performance of the four existing methods, **HHC1** and **HHCL** (Huang et al., 2011), **band** (Wu and Pourahmadi, 2009), **spec** (Fan and Kreutzberger, 1998) with the

proposed estimators with  $\alpha = q = 2$ , **GCV**, **GCVM** (modified GCV), **BIC**, **L-curve**, **L-curveEM** (L-curve with varying  $0 < q \leq 1$ ). For **HHC1**, **HHCL** and the proposed methods, we let  $\lambda$  range from 1 to 10000 for the MA cases and from 0.0001 to 100 for the AR cases with the gap,  $\log(\lambda_{i+1}) - \log(\lambda_i) = 0.1$ . For **L-curveEM**, we try  $q = 0.1, 0.2, \dots, 1$ .

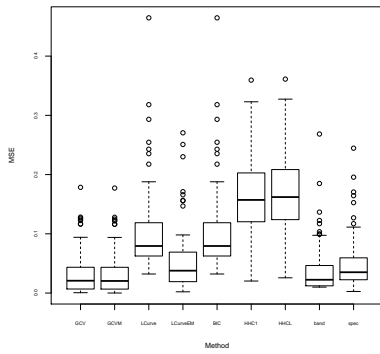
## 4.1 Comparison of Mean Squared Errors

Figures 1 and 2 display the boxplots of MSE values for testing data sets. We do not report the results from training data sets to save space, but report them in the supplementary materials. We draw the boxplots without **BIC** in Figures 2(a)–(c) because of its poor performance. We also exclude **GCV** in Figures 2(c) for the same reason.

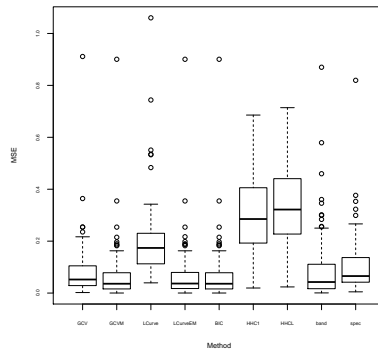
For the MA cases in Figure 1, the proposed estimators **GCVM** and **L-curveEM** consistently outperform the other methods for the MA(1) models with zero and nonzero means. For MA(2), **L-curveEM** yields the smallest MSE values for both zero and nonzero models. It is noticeable that the other methods do not perform well for nonzero mean cases, which might be caused by the detrending procedure we apply for this simulation. Since the proposed method is based on differenced time series, it does not rely on a particular procedure of estimating a deterministic function  $g$ .

For the AR cases with zero mean in Figure 2, **GCVM** performs best for weak (coefficient 0.1) and moderate correlations (coefficient 0.5) while **spec** shows the best performance for strong correlation (coefficient 0.9, and AR(2)). This is somewhat expected because the idea of using a penalty in the estimation is based on the assumption that serial correlation in a time series is small at large time lags. Among the proposed estimators, two L-curve approaches perform better than the GCV and BIC approaches for the strong correlation case. For the nonzero mean time series, it is again observed that the proposed methods tend to outperform the existing methods.

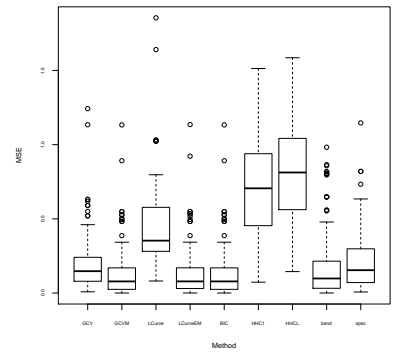
Figures 3 and 4 display the 100 averaged estimates by the four proposed methods along with the true function for each model. For all the MA models and the AR models with weak or moderate serial correlation, it can be seen that the differences among the proposed methods are minimal. For AR(1) with coefficient 0.9, it is evidenced that the estimates of the L-curve approach are closer to the true function than those of the GCV and BIC approaches. The **GCV** and **BIC** yield the unstable estimates for large time lags, and **GCVM** does not catch the decaying slope correctly for small time lags. For the nonzero mean AR(1) with coefficient 0.9 and AR(2) with both zero and nonzero means, all the proposed methods do



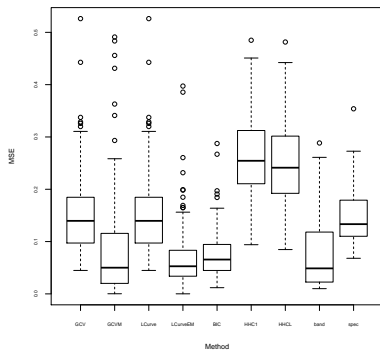
(a) MA(1): 0.1



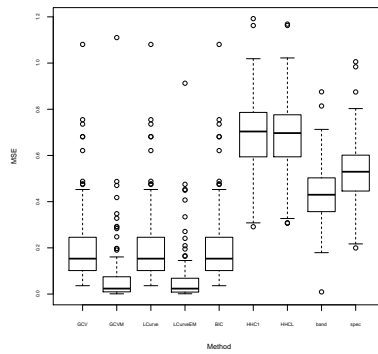
(b) MA(1): 0.5



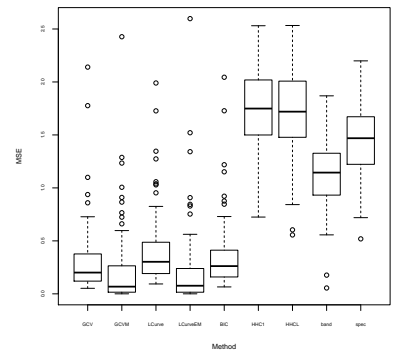
(c) MA(1): 0.9



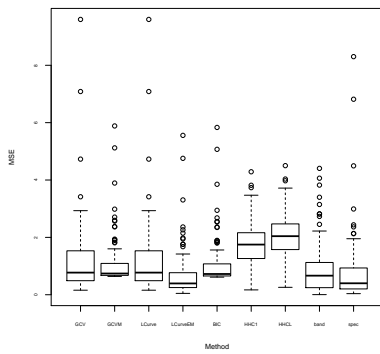
(d)  $g+MA(1)$ : 0.1



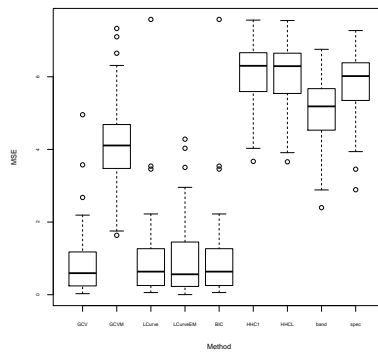
(e)  $g+MA(1)$ : 0.5



(f)  $g+MA(1)$ : 0.9

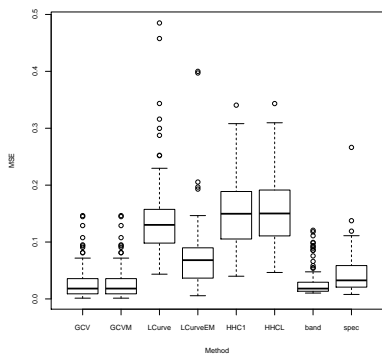


(g) MA(2): (0.9,0.8)

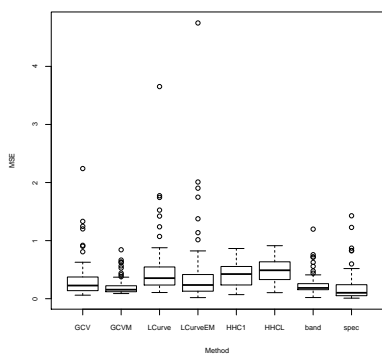


(h)  $g+MA(2)$ : (0.9,0.8)

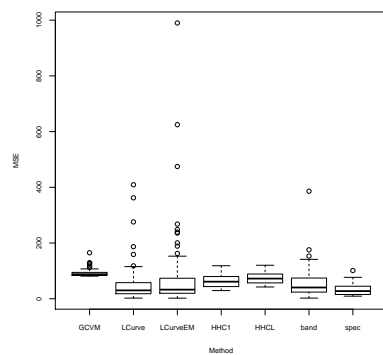
Figure 1: Boxplots of Average MSE for MA cases



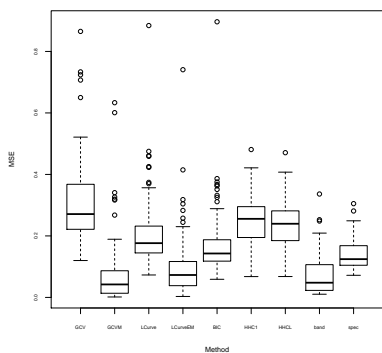
(a) AR(1): 0.1



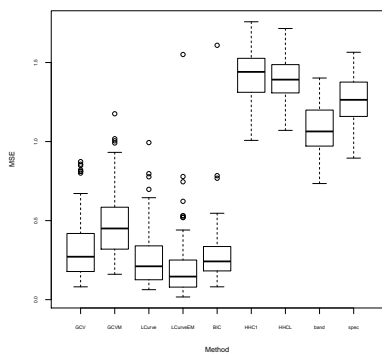
(b) AR(1): 0.5



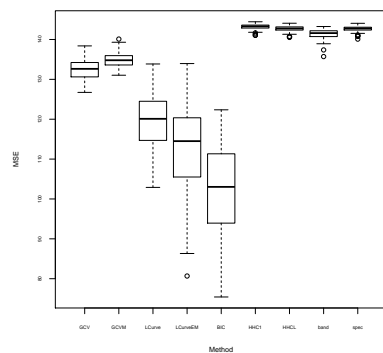
(c) AR(1): 0.9



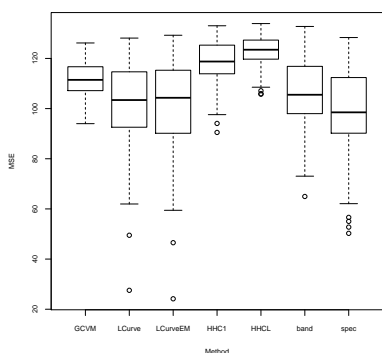
(d)  $g + \text{AR}(1)$ : 0.1



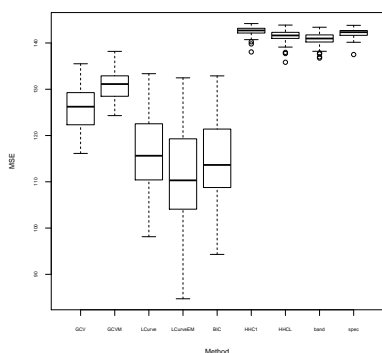
(e)  $g + \text{AR}(1)$ : 0.5



(f)  $g + \text{AR}(1)$ : 0.9

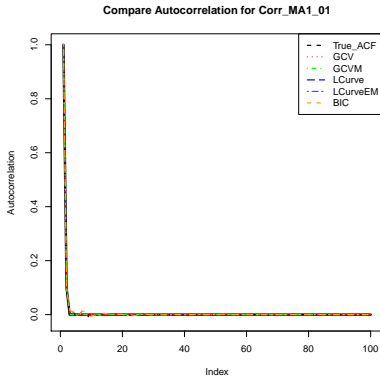


(g) AR(2):  $(-0.9, 0.2)$

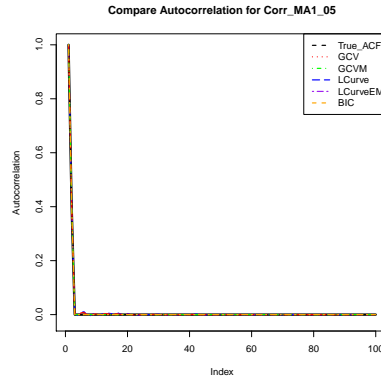


(h)  $g + \text{AR}(2)$ :  $(-0.9, 0.2)$

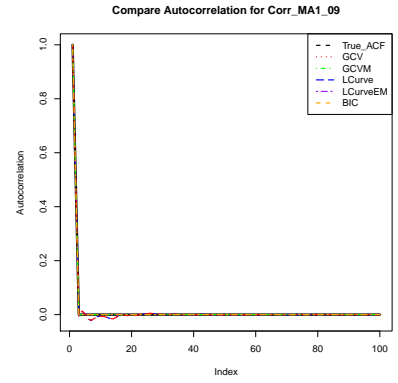
Figure 2: Boxplots of Average MSE for AR cases. We exclude BIC for (a) and (b), and GCV and BIC for (c) due to their poor performance.



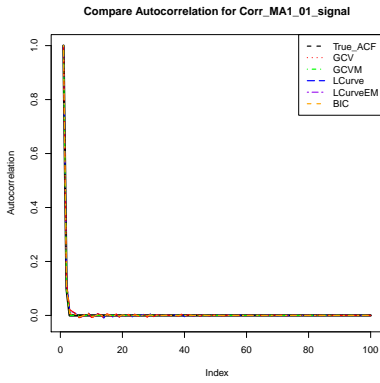
(a) MA(1): 0.1



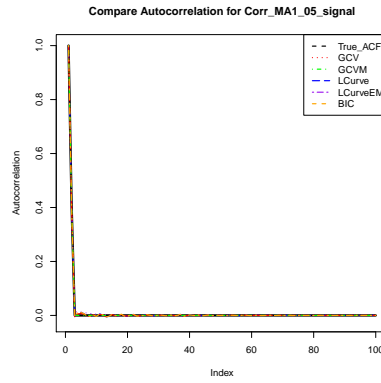
(b) MA(1): 0.5



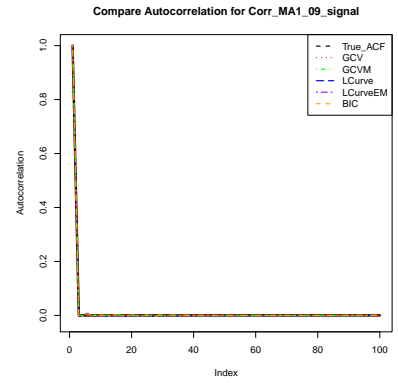
(c) MA(1): 0.9



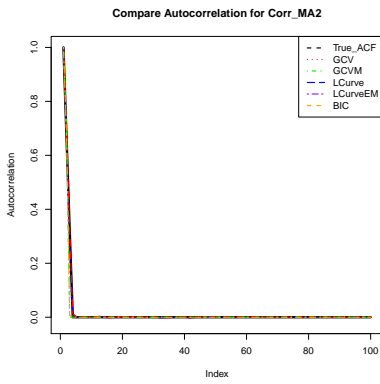
(d)  $g+MA(1)$ : 0.1



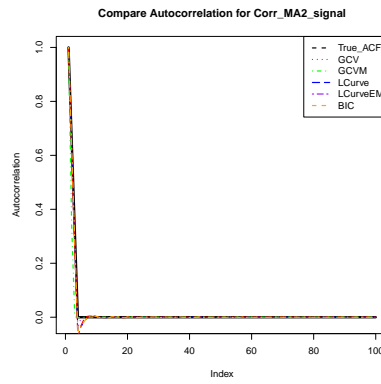
(e)  $g+MA(1)$ : 0.5



(f)  $g+MA(1)$ : 0.9

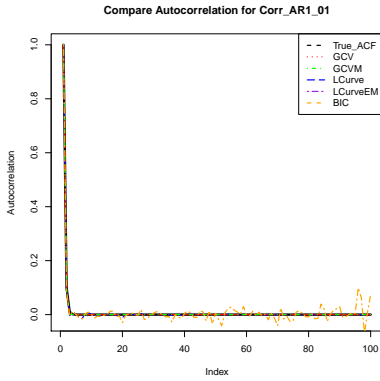


(g) MA(2): (0.9,0.8)

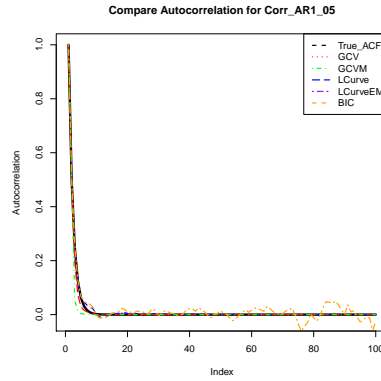


(h)  $g+MA(2)$ : (0.9,0.8)

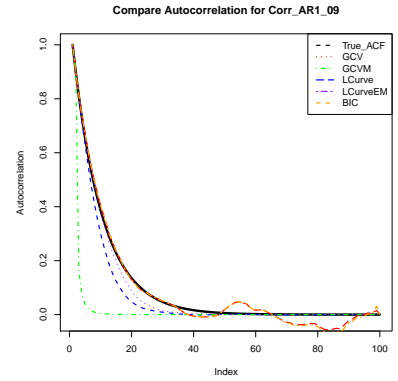
Figure 3: Autocorrelation Function for MA Cases



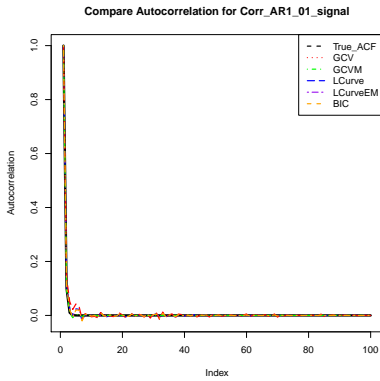
(a) AR(1): 0.1



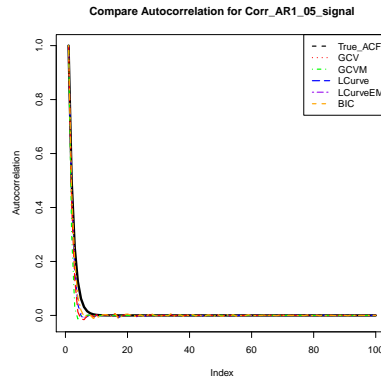
(b) AR(1): 0.5



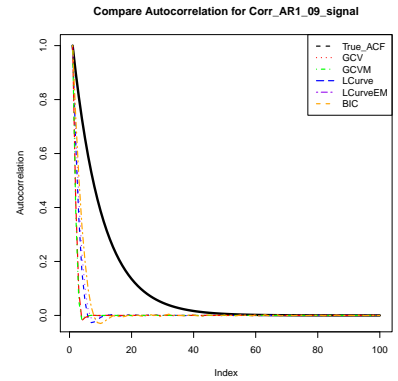
(c) AR(1): 0.9



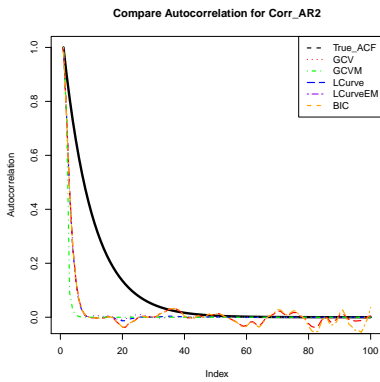
(d)  $g + \text{AR}(1)$ : 0.1



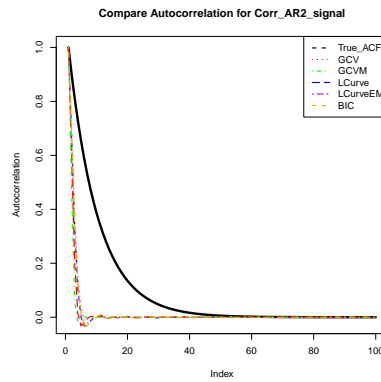
(e)  $g + \text{AR}(1)$ : 0.5



(f)  $g + \text{AR}(1)$ : 0.9



(g) AR(2): (-0.9, 0.2)



(h)  $g + \text{AR}(2)$ : (-0.9, 0.2)

Figure 4: Autocorrelation function for AR cases

not succeed in estimating the true function closely for small time lags although the L-curve approach performs slightly better. This shows the limitation of the proposed approach for strongly correlated time series.

To sum up, this simulation study demonstrates that the proposed estimators outperform existing methods in most cases, and show the competitive performance for strongly correlated time series. For nonzero mean time series, the proposed methods, which remove the effect of the mean function  $g$  by taking differences on a time series, show superior performance to the existing methods, which detrend the mean function by separately estimating it. Among the proposed methods, it is not straightforward to single out the best one, but we do not recommend to use the GCV and BIC approaches due to its poor performance for some cases.

## 4.2 Nonnegative Definiteness

One of the key properties of covariance function is nonnegative definiteness. To explore this property in the proposed methods, we investigate eigenvalues of the estimated covariance matrix from the simulated examples.

Table 1: Average number of negative eigenvalues with standard error in brackets for all methods for zero mean MA cases

Class	MA(1): 0.1	MA(1): 0.5	MA(1): 0.9	MA(2)
GCV	0 (0)	0 (0)	1.87 (0.29)	0.16 (0.07)
GCVM	0 (0)	1.18 (0.36)	4.93 (0.66)	22.3 (0.34)
L-curve	0 (0)	0 (0)	0.89 (0.18)	0.16 (0.07)
L-curveEM	0 (0)	0.85 (0.28)	4.41 (0.59)	14.6 (0.97)
BIC	0 (0)	1.12 (0.35)	4.91 (0.65)	22.2 (0.35)
HHC1	0 (0)	0 (0)	0 (0)	0 (0)
HHCL	0 (0)	0 (0)	0 (0)	0 (0)
band	0 (0)	0.73 (0.24)	4.45 (0.60)	11.5 (0.83)
spec	0 (0)	0 (0)	0 (0)	0 (0)

Tables 1–4 report the average number of negative eigenvalues with its standard error in brackets for all methods and for each simulation setting. Since **HHC1**, **HHCL**, and **spec** guarantee the nonnegative definiteness, they do not have any negative values in the tables. It



Table 2: Average number of negative eigenvalues with standard error in brackets for all methods for nonzero mean MA cases

Class	$g + \text{MA}(1): 0.1$	$g + \text{MA}(1): 0.5$	$g + \text{MA}(1): 0.9$	$g + \text{MA}(2)$
GCV	0 (0)	0.01 (0.01)	0.73 (0.12)	0 (0)
GCVM	0.90 (0.45)	0 (0)	3.31 (0.39)	0.04 (0.04)
L-curve	0 (0)	0.01 (0.01)	0.82 (0.12)	0.06 (0.03)
L-curveEM	0 (0)	0 (0)	3.25 (0.38)	0.87 (0.26)
BIC	0 (0)	0.01 (0.01)	0.74 (0.12)	0.06 (0.03)
HHC1	0 (0)	0 (0)	0 (0)	0 (0)
HHCL	0 (0)	0 (0)	0 (0)	0 (0)
band	0.04 (0.04)	1.20 (0.22)	5.01 (0.46)	7.50 (0.62)
spec	0 (0)	0 (0)	0 (0)	0 (0)

Table 3: Average number of negative eigenvalues with standard error in brackets for all methods for zero mean AR cases

Class	$AR(1) : 0.1$	$AR(1) : 0.5$	$AR(1) : 0.9$	$AR(2)$
GCV	0 (0)	0 (0)	24.5 (0.45)	12.0 (0.38)
GCVM	0 (0)	3.45 (0.65)	31.6 (0.15)	24.9 (0.41)
L-curve	0 (0)	0 (0)	0.41 (0.17)	0 (0)
L-curveEM	0 (0)	0.14 (0.06)	2.79 (0.34)	2.91 (0.29)
BIC	20.2 (0.34)	20.6 (0.31)	23.8 (0.56)	20.8 (0.44)
HHC1	0 (0)	0 (0)	0 (0)	0 (0)
HHCL	0 (0)	0 (0)	0 (0)	0 (0)
band	0 (0)	1.38 (0.38)	1.76 (0.28)	4.48 (0.75)
spec	0 (0)	0 (0)	0 (0)	0 (0)

is evident that the proposed estimators do not guarantee this property if serial correlation is moderate or strong. It can be also seen that **L-curve** produces a smaller number of negative eigenvalues for all cases while **GCV**, **GCVM**, and **BIC** tend to produce a large number for  $AR(1)$  with 0.9 and  $AR(2)$ . It might suggest that selection of the tuning parameter  $\lambda$  could be designed in a way of warranting the nonnegative definiteness. We leave this issue as our

Table 4: Average number of negative eigenvalues with standard error in brackets for all methods for nonzero mean AR cases

Class	$g + AR(1) : 0.1$	$g + AR(1) : 0.5$	$g + AR(1) : 0.9$	$g + AR(2)$
GCV	0.39 (0.09)	0 (0)	0 (0)	0 (0)
GCVM	0 (0)	0 (0)	0 (0)	0 (0)
L-curve	0.10 (0.04)	0.08 (0.04)	0.02 (0.02)	0.07 (0.03)
L-curveEM	0.08 (0.05)	0.13 (0.06)	0.09 (0.05)	0.27 (0.13)
BIC	0.01 (0.01)	0.32 (0.07)	2.51 (0.18)	0.22 (0.06)
HHC1	0 (0)	0 (0)	0 (0)	0(0)
HHCL	0 (0)	0 (0)	0 (0)	0(0)
band	0.08 (0.05)	0.60 (0.15)	0.35 (0.03)	1.24 (0.21)
spec	0 (0)	0 (0)	0 (0)	0 (0)

future study.

As described in Section 3.1, a simple fix to negative eigenvalues of the estimate is to replace them with zero values and revise the estimate. Tables 5 and 6 compare the original MSE values with revised ones for MA(1) and AR(1) with coefficient 0.9, respectively, where there are significant numbers of negative eigenvalues in the original estimates. It can be seen that the differences between the two estimates are minimal in all cases for proposed methods. Therefore, we suggest the revised estimate by removing negative eigenvalues as a practical solution.

## 5 Discussion

In this section, we discuss possible extension of the proposed work to the estimation of an autocovariance function in interval-valued time series. Data collection has become easier and more widespread over the last decade due to technological advances and aggregating the information in huge databases is routinely done. A current challenge is to extract knowledge from large data bases by summarizing these data into new meaningful formats. This process creates a new type of data that defies traditional statistical approaches. Symbolic data (Billard and Diday, 2007) including intervals, lists and histograms are such examples. They are more complex than other data formats because they contain internal variation and are

structured. Interval-valued time series are interval-valued data collected in a chronological sequence. For instance, the daily or weekly highest and lowest prices of assets may be regarded as boundary values of an interval. In economic and financial settings, since data such as stocks, exchange rate, oil price, and electricity demand are observed in a fine grid, they are often summarized as interval-valued time series to reflect variation information. While a few forecasting models have been developed for interval-valued time series (e.g. Maia et al. (2008)), the fundamental problem of estimating serial correlation has not been studied to our best knowledge. We suggest the estimation of an autocovariance function for interval-valued time series as our future work.

## 6 Appendix

### 6.1 Derivation of the Estimator

For the special case of  $\alpha = 2$  and  $q = 2$ ,  $f_n(\boldsymbol{\gamma}; \lambda)$  in (3.1) can be rewritten in the matrix form as follows:

$$f_n(\boldsymbol{\gamma}; \lambda) = (\mathbf{y} - \mathbf{X}\boldsymbol{\gamma})^\top(\mathbf{y} - \mathbf{X}\boldsymbol{\gamma}) + \lambda(\mathbf{A}\boldsymbol{\gamma})^\top(\mathbf{A}\boldsymbol{\gamma}).$$

The first derivative of  $f_n(\boldsymbol{\gamma}; \lambda)$  with respect to  $\gamma(k)$  are

$$\begin{aligned} \frac{\partial f_n(\boldsymbol{\gamma}; \lambda)}{\partial \gamma(0)} &= 2n\gamma(0) - 2\mathbf{z}_0^\top \mathbf{1}_n \\ \frac{\partial f_n(\boldsymbol{\gamma}; \lambda)}{\partial \gamma(k)} &= 2(n-k)\gamma(k) - 2\mathbf{z}_k^\top \mathbf{1}_{n-k} + 2\lambda k^2 \gamma(k). \end{aligned}$$

Let  $\frac{\partial f_n(\boldsymbol{\gamma}; \lambda)}{\partial \gamma(k)} = 0$  for  $k = 0, 1, \dots, n-1$ . Solving these equations gives

$$\hat{\gamma}(0) = \frac{1}{n} \mathbf{1}_n^\top \mathbf{z}_0, \quad \hat{\gamma}(k) = \frac{1}{(n-k) + \lambda k^2} \mathbf{1}_{n-k}^\top \mathbf{z}_k, \quad \text{for } k = 1, \dots, n-1.$$

We note that if  $k \leq \lambda k^2$ ,

$$|\hat{\gamma}(k)| \leq |\bar{\gamma}(k)| \leq \bar{\gamma}(0) = \hat{\gamma}(0).$$

Hence, the proposed estimator satisfies the restriction  $\mathcal{C} = \{\gamma : \gamma(0) \geq |\gamma(k)|, k = 1, \dots, n-1\}$ . Here, we use the fact that  $\bar{\gamma}(k)$ , the sample autocovariance function defined in (1.2), is nonnegative definite, and thus  $|\bar{\gamma}(k)| \leq \bar{\gamma}(0)$ .

## 6.2 Proofs of Theorems 1 and 2

We present the proofs of Theorems 1 and 2 by following the setting and notation in Wu (2009). Let  $\kappa_p = \|Y_i\|_p$ .

For Theorem 1, we use the coupling argument and central limit theorem for stationary processes (Hannan, 1973). Similar calculations to the proof of Theorem 1 in Wu (2009) give

$$\sum_{i=0}^{\infty} \|\mathcal{P}_0(Y_i Y_{i-k})\| \leq 2\kappa_4 \sum_{i=0}^{\infty} \delta_4(i) < \infty.$$

Also, for  $0 \leq j \leq k$  and appropriate  $\lambda$ ,  $\frac{n}{(n-j)+\lambda j^2} \leq 1$ , which results in  $\sum_{i=0}^{\infty} \|\mathcal{P}_0(Y_i \mathbf{A}_k \boldsymbol{\eta}_i)\| < \infty$ . Hence, by the Cramer-Wold device, (3.4) follows from Theorem 1 in (Hannan, 1973).

For Theorem 2, let  $T_{nj} = \sum_{i=1}^{n-j} (Y_i Y_{i+j} - \gamma(j))$ . The calculations of the proof are similar to the proof of Theorem 7 (i) in Wu (2011). In particular following Wu (2011) we obtain:

$$\|\mathcal{P}_0(Y_i Y_{i+j})\|_{p/2} \leq \|Y_i\|_p \delta_p(i+j) + \delta_p(i) \|Y'_{i+j}\|_p,$$

for  $j \geq 0$ . By the triangle inequality,

$$\|T_{nj}\|_{p/2} = \left\| \sum_{i=1}^{n-j} \sum_{l \in \mathbb{Z}} \mathcal{P}_{i-l} Y_i Y_{i+j} \right\|_{p/2} \leq \sum_{l \in \mathbb{Z}} \left\| \sum_{i=1}^{n-j} \mathcal{P}_{i-l} Y_i Y_{i+j} \right\|_{p/2}.$$

Note that  $\mathcal{P}_{i-l} Y_i Y_{i+j}$ ,  $i = 1, \dots, n$ , form stationary martingale differences. Using Burkholder's moment inequality for martingale differences (Burkholder, 1988), the fact  $p/4 \leq 1$  and simple algebra we obtain

$$\left\| \sum_{i=1}^{n-j} \mathcal{P}_{i-l} Y_i Y_{i+j} \right\|_{p/2} \leq \frac{\mathbb{E}\{[\sum_{i=1}^{n-j} (\mathcal{P}_{i-l} Y_i Y_{i+j})^2]^{p/4}\}}{(p/2 - 1)^{p/2}} \leq \frac{(n-j) \|\mathcal{P}_0 Y_l Y_{l+j}\|_{p/2}^{p/2}}{(p/2 - 1)^{p/2}}.$$

From these three facts

$$\begin{aligned} \|T_{nj}\|_{p/2} &\leq \sum_{l \in \mathbb{Z}} \left\| \sum_{i=1}^{n-j} \mathcal{P}_{i-l} Y_i Y_{i+j} \right\|_{p/2} \\ &\leq \sum_{l \in \mathbb{Z}} \frac{(n-j)^{2/p} \|\mathcal{P}_0 Y_l Y_{l+j}\|_{p/2}}{p/2 - 1} \\ &= \frac{(n-j)^{2/p}}{p/2 - 1} \sum_{l \in \mathbb{Z}} \left( \|Y_l\|_p \delta_p(l+j) + \delta_p(l) \|Y'_{l+j}\|_p \right) \\ &\leq \frac{4(n-j)^{2/p} \kappa_p \Delta_p}{p-2}. \end{aligned}$$

Since  $\delta_p(i) = 0$  if  $i < 0$ ,

$$\begin{aligned} \hat{\gamma}(k) - \frac{n-k}{(n-k) + \lambda k^2} \gamma(k) &= \frac{1}{(n-k) + \lambda k^2} \sum_{i=1}^{n-k} (Y_i Y_{i+k} - \gamma(k)) \\ &\leq \frac{1}{n-k} \sum_{i=1}^{n-k} (Y_i Y_{i-k} - \gamma(k)) \\ &= \frac{1}{n-k} T_{nk} \end{aligned}$$

for appropriate  $\lambda$ . Hence,

$$\begin{aligned} \left\| \hat{\gamma}(k) - \left( 1 - \frac{\lambda k^2}{(n-k) + \lambda k^2} \right) \gamma(k) \right\|_{p/2} &\leq \frac{4(n-k)^{2/p} \kappa_p \Delta_p}{(p-2)n} \\ &= \frac{4 \|Y_1\|_p \Delta_p}{(p-2)n^{1-2/p}}. \end{aligned}$$

## References

- Billard, L. and Diday, E. (2007). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley, Chichester.
- Burkholder, D. L. (1988). Sharp inequalities for martingales and stochastic integrals. In *Colloque Paul Lévy sur les Processus Stochastiques*, pages 75–94. Ecole polytechnique, Palaiseau. Astérisque No. 157-158.
- Fan, J. and Kreutzberger, E. (1998). Automatic local smoothing for spectral density estimation. *Scandinavian Journal of Statistics*, 25:359–369.
- Hall, P., Fisher, N. I., and Hoffmann, B. (1994). On the nonparametric estimation of covariance functions. *The Annals of Statistics*, 22:2115–2134.
- Hannan, E. J. (1973). Central limit theorems for time series regression. *Z. Wahrsch. Verw. Gebiete*, 26:157–170.
- Hansen, P. C. (1987). *Rank-Deficient and Discrete Ill-Posed Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- Hansen, P. C. (2000). The L-curve and its use in the numerical treatment of inverse problems. In *Computational Inverse Problems in Electrocardiology*, ed. P. Johnston, *Advances in Computational Bioengineering*, pages 119–142.

- Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12:55–67.
- Huang, C., Hsing, T., and Cressie, N. (2011). Spectral density estimation through a regularized inverse problem. *Statistica Sinica*, 21:1115–1144.
- Hyndman, R. J. and Wand, M. P. (1997). Nonparametric autocovariance function estimation. *Australian Journal of Statistics*, 39:313–325.
- Maia, A. L. S., de A.T. de Carvalho, F., and Ludermir, T. B. (2008). Forecasting models for interval-valued time series. *Neurocomputing*, 71:3344–3352.
- Park, C., Hannig, J., and Kang, K. (2009). Improved sizer for time series. *Statistica Sinica*, 19:1511–1530.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.
- Whittle, P. (1962). Gaussian estimation in stationary time series. *Bulletin of the International Statistical Institute*, 39:105–129.
- Wu, W. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Journal Acad. Sci.*, 102(40):14150–14154.
- Wu, W. B. (2009). An asymptotic theory for sample covariances of bernoulli shifts. *Stochastic Processes and their Applications*, 119:453–467.
- Wu, W. B. (2011). Asymptotic theory for stationary processes. *Statistics and its Interface*, 4(2):207–226.
- Wu, W. B. and Pourahmadi, M. (2009). Banding sample autocovariance matrices of stationary processes. *Statistica Sinica*, 19:1755–1768.
- Zou, H., Hastie, T., and Tibshirani, R. (2007). On the degrees of freedom of the lasso. 35:2173–2192.

Table 5: Comparison of average MSE and standard error for MA(1): 0.9

Class	Zero		Nonzero	
	Original	Revised	Original	Revised
GCV	0.21 (0.021)	0.21 (0.021)	0.30 (0.031)	0.30 (0.031)
GCVM	0.14 (0.019)	0.14 (0.019)	0.21 (0.035)	0.21 (0.035)
L-curve	0.44 (0.029)	0.44 (0.029)	0.40 (0.033)	0.40 (0.033)
L-curveEM	0.15 (0.019)	0.15 (0.019)	0.20 (0.036)	0.20 (0.036)
BIC	0.14 (0.019)	0.14 (0.019)	0.35 (0.032)	0.35 (0.032)

Table 6: Comparison of average MSE and standard error for AR(1): 0.9

Class	Zero		Nonzero	
	Original	Revised	Original	Revised
GCV	469 (57.9)	437 (56.7)	132 (0.27)	132 (0.27)
GCVM	91.3 (1.22)	90.2 (1.18)	135 (0.19)	135 (0.19)
L-curve	50.0 (6.34)	50.0 (6.34)	119 (0.68)	119 (0.68)
L-curveEM	74.4 (12.9)	74.4 (12.9)	113 (1.07)	113 (1.07)
BIC	541 (67.2)	487 (64.3)	102 (1.09)	102 (1.09)