Statistical inference on $AR(p)$ models with non-i.i.d. innovations

Yunyi Zhang and Dimitris N. Politis

October 28, 2021

Abstract

The autoregressive process is one of the fundamental and most important models that analyze a time series. Theoretical results and practical tools for fitting an autoregressive process with i.i.d. innovations are well-established. However, when the innovations are white noise but not i.i.d., those tools fail to generate a consistent confidence interval for the autoregressive coefficients. Focus on an autoregressive process with dependent and non-stationary innovations, this paper provides a consistent result and a Gaussian approximation theorem for the Yule-Walker estimator. Moreover, it introduces the second order wild bootstrap that constructs a consistent confidence interval for the estimator. Numerical experiments confirm the validity of the proposed algorithm with different kinds of white noise innovations. Meanwhile, the classical method (e.g., AR(Sieve) bootstrap) fails to generate a correct confidence interval when the innovations are dependent.

According to Kreiss et al. [1] and the Wold decomposition, assuming a real-life time series satisfies an autoregressive process is reasonable. However, innovations in that process are more likely to be white noises instead of i.i.d.. Therefore, our method should provide a practical tool that handles real-life problems.
1 Introduction

The autoregressive process with lag $p$ (AR($p$))

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \epsilon_t, \ t \in \mathbb{Z}, \ \epsilon_t \sim WN(0, \sigma^2)$$

$(1)$

$WN$ stands for ‘white noise’ and $\epsilon_t$ is called innovations

is one of the fundamental and most important models for modelling a time series. The reason mainly comes from the following 4 aspects:

1. Fitting an AR($p$) model is straightforward. The Yule-Walker equation (e.g., section 8.1 of Brockwell and Davis [2]) is a classical method that solves this problem. Besides, an AR($p$) model assumes a linear relationship between $X_t$ and $\{X_{t-1}, ..., X_{t-p}\}$. Therefore, linear regression algorithms (e.g., Lasso or weighted least squares, see Nardi and Rinaldo [3] and Li and Politis [4]) can be applied to fit the model. We also refer Han et al. [5], Basu and Michalidis [6] and Krampe and Paparoditis [7] as an overview for fitting a multivariate (high-dimensional) autoregressive process. Meanwhile, fitting a general time series is more difficult. Fan and Yao [8] and Li and Politis [4] applied the Kernel estimator; Kirch and Kamgaing [9] considered a neural network and Cai et al. [10] applied the functional-coefficient autoregressive model to fit a non-linear time series. Notably, even if a time series is linear, it still can be hard to fit. For example, fitting a MA (moving average) process requires the innovation algorithm, and fitting an ARMA (autoregressive and moving average) process relies on the maximum likelihood estimation, see section 8.3 in Brockwell and Davis [2]. Compared to the Yule-Walker equation, the innovation algorithm estimator does not have a closed-form formula, making it hard to derive the theoretical guarantees.

2. AR($p$) model approximates a wide range of (nonlinear) time series. Suppose $X_i, i \in \mathbb{Z}$ is a stationary time series. According to Kreiss et al. [11], under some assumptions there exists a sequence of coefficients $\{a_i\}_{i=1,2,...}$ such that

$$X_i = \sum_{j=1}^{\infty} a_j X_{i-j} + \epsilon_i, \ \epsilon_i \sim WN(0, \sigma^2)$$

$(2)$

so the time series $X_i$ can be expressed as an AR($\infty$) process. This result is a counterpart of the famous Wold decomposition (theorem 5.7.1 in Brockwell and Davis [2]) that expresses $X_i$ as a MA($\infty$) process.
3. Even if a stationary mean 0 time series $X_i$ cannot be expressed as an $AR(p)$ process, fitting an $AR(p)$ model on $X_i$ is still beneficial. For a fixed index $i$, define the Hilbert space $H$ spanned by $X_{i-1}, X_{i-2}, \ldots, X_{i-p}$ with the inner product $\langle X, Y \rangle = E(XY)$. Then the projection of $X_i$ on $H$ will be $\sum_{j=1}^{p} a_j X_{i-j}$. Here $(a_1', \ldots, a_p')^T$ satisfies the Yule-Walker equation $\Gamma(a_1', \ldots, a_p')^T = (\gamma_1, \ldots, \gamma_p)^T$, $\Gamma = (\gamma_{|i-j|})_{i,j=1,\ldots,p}$, $\gamma_i = EX_0 X_i$, see section 2 in Brockwell and Davis [2] and section 3.2 in Fan and Yao [10]. Therefore, fitting an $AR(p)$ model on $X_i$ contributes to the optimal linear predictor for future observations.

4. Statisticians can use a bunch of $AR(p)$ processes (with different autoregressive coefficients) to model a nonlinear / non-stationary time series. This idea leads to the ‘threshold autoregressive model’ (section 4.1.1 in Fan and Yao [8]) and the ‘functional coefficient autoregressive model’ (Chen and Tsay [11], Cai et al. [10] and Xia and Li [12]).

Despite its simplicity and universality, statistical inference on an $AR(p)$ process is a big topic. Section 8.10 in Brockwell and Davis [2] derived the asymptotic normality of the Yule-Walker estimator; Kreiss [13] considered the AR bootstrap and Bühlmann [14] adjusted the AR bootstrap to an $AR(\infty)$ model. We also refer Paparoditis [15] for bootstrapping a vector autoregressive model and Braumann et al. [16] for a simultaneous inference on autocovariance. A surprising fact is that the Yule-Walker estimator’s asymptotic covariance matrix only depends on the second order structures, i.e., the innovation’s variance and the autocovariance of the time series. However, these results rely on the fact that the innovation $\epsilon_t, t \in \mathbb{Z}$ in (1) are i.i.d.

This paper starts with the question

What happens if the innovations in model (1) are white noise, but dependent (or even worse, non-stationary)?

This question is meaningful for the following reasons: one the one hand, a general (nonlinear) time series may still have $AR(\infty)$ or $MA(\infty)$ expression according to Kreiss et al. [11] and the Wold decomposition. So statisticians can truncate the corresponding $AR(\infty)$ expression and use an $AR(p)$ (with sufficiently large $p$) process to model the time series. Nevertheless, innovations in those expressions are white noise rather than i.i.d.. On the other hand, whether or not the innovations are white noise can be tested by methods introduced in Li et al. [17], Bagchi et al. [18] and their reference. But it is harder to test the innovations being i.i.d..

Example 1
Suppose \( e_i, i \in \mathbb{Z} \) are i.i.d. normal random variables with mean 0 and variance 1. Suppose 3 types of innovations: 1. normal: \( \epsilon_i = e_i \); 2. product normal: \( \epsilon_i = e_i e_{i-1} \); 3. non-stationary: \( \epsilon_i = e_i \) and \( \epsilon_{i+1} = e_{i+1} e_{i-1} \). All of them are white noise, i.e., \( \mathbb{E} \epsilon_i = 0 \), \( \mathbb{E} \epsilon_i \epsilon_j = 0 \) for \( i \neq j \) and \( \mathbb{E} \epsilon_i^2 = 1 \). However, for case 2 \( \mathbb{E} \epsilon_i^2 \epsilon_{i-1}^2 = \mathbb{E} e_i^2 \times \mathbb{E} e_{i-2}^2 \times \mathbb{E} e_{i-1}^2 = 3 \neq 1 \), so innovations in case 2 are dependent. Moreover, innovations in case 3 are non-stationary.

Define \( X_i = \epsilon_i + \sum_{j=1}^{\infty} \rho^j \epsilon_{i-j}, i \in \mathbb{Z} \), then \( X_i = \rho X_{i-1} + \epsilon_i \), i.e., \( X_i \) is an AR(1) process. Here \( |\rho| < 1 \), \( \mathbb{E} X_i = 0 \) and \( \mathbb{E} X_i X_{i+k} = \frac{\rho^k}{1-\rho^2}, k \geq 0 \) for all innovations. Table 1 calculates the variance of the Yule-Walker estimator \( \hat{\rho} \) as well as the sample autocovariance \( \hat{\gamma}_1 = \frac{1}{n} \sum_{i=2}^{n} X_i X_{i-1} \). Case 1 coincides with theorem 8.1.1 in Brockwell and Davis [2], i.e., the asymptotic variance of \( \hat{\rho} \) is \( 1 - \rho^2 \). However, despite the estimators’ consistency is assured in all cases, in case 2 and 3 the second order structures cannot decide the variance of \( \hat{\rho} \) and \( \hat{\gamma}_1 \). Notably, this phenomenon does not have conflict with Braumann et al. [16] and Xiao and Wu [19], for their works control \( \sqrt{n} \max_{1 \leq i \leq s} |\hat{\gamma}_i - \gamma_i| (\hat{\gamma}_k \text{ is the } k\text{-th sample autocovariance}) \) and require \( s \to \infty \); while our work focuses on some specific sample autocovariance.

| Residual type   | n    | \( \rho \) | \( \text{Var}(\sqrt{n} (\hat{\rho} - \rho)) \) | \( \hat{\gamma}_1 \) | \( \text{Var}(\sqrt{n} (\hat{\gamma}_1 - \gamma_1)) \) |
|-----------------|------|------------|---------------------------------|----------------|---------------------------------|
| Normal          | 10000| 0.70       | 0.52                            | 1.45           | 20.74                           |
|                 | 50000| 0.70       | 0.51                            | 1.39           | 20.83                           |
|                 | 100000| 0.70      | 0.52                            | 1.37           | 21.00                           |
| Product normal  | 10000| 0.69       | 1.03                            | 1.31           | 57.34                           |
|                 | 50000| 0.70       | 1.02                            | 1.40           | 57.78                           |
|                 | 100000| 0.70      | 1.04                            | 1.40           | 57.54                           |
| Non-stationary  | 10000| 0.71       | 1.58                            | 1.49           | 70.51                           |
|                 | 50000| 0.70       | 1.55                            | 1.42           | 72.18                           |
|                 | 100000| 0.70     | 1.53                            | 1.40           | 70.68                           |

If innovations in a time series are not i.i.d., example 1 shows that the Yule-Walker estimator’s variance no longer coincides with the i.i.d. situation. Therefore, the validity of other results, including the central limit theorem and the bootstrap, is not obvious. To complement the theoretical results, Dette et al. [20] derived the central limit theorem for the variance estimator of a piece-wise stationary time series; Jiang and Politis [21] constructed an AR-based spectral density estimator for nonlinear time series and proved its consistency; Zhang and Wu [22] derived oracle inequalities for the spectral density estimator of a high dimensional time series; McMurry and Politis [23], Jentsch and Politis [24], Kreiss et al. [1]...
and Fragkeskou and Paparoditis [25] discussed the validity of different kinds of bootstrap algorithms. Based on their discussion, the consistency of the AR(Sieve) bootstrap is assured if the estimator’s asymptotic variance only depends on the second order structures of the time series. Otherwise, the AR(Sieve) bootstrap approximates the distribution of a ‘companion’ autoregressive process (see Kreiss et al. [1]), i.e., the linear process whose second order structures coincide with the original time series but has i.i.d. innovations. Many important estimators, including the sample mean, the Xiao and Wu’s test statistics [19] and the sample spectral density, suit this case. However, –as example 1 demonstrates–, second order structures of a time series is not sufficient to determine the Yule-Walker estimator’s variance. So the AR bootstrap fails to provide a consistent confidence interval.

This paper aims at providing a consistent confidence interval for the Yule-Walker estimator. Focus on an $AR(p)$ process with white noise (non-i.i.d.) innovations, we establish the Gaussian approximation theorem for the Yule-Walker estimator. Moreover, we provide a bootstrap algorithm (named second-order wild bootstrap) that automatically generates the consistent simultaneous confidence intervals for the estimator. Notably, our work does not require the ‘strict stationary’ assumptions. So our work is able to handle the $AR(p)$ process with non-stationary white noise innovations like case 3 in example 1.

The remaining parts of this paper are as follows: section 2 introduces the frequently used notations and the basic assumptions for this paper. Section 3 presents some properties of a non-stationary time series. Section 4 derives the consistency and the Gaussian approximation theorem for the Yule-Walker estimator. Section 5 introduces the second order wild bootstrap algorithm for making confidence intervals / testing statistical hypothesis. Section 6 provides some numerical experiments that illustrate the finite sample performance of our algorithm. Section 7 makes the conclusion. The theoretical proofs will be postponed to an appendix.

## 2 Preliminary

Suppose random variables $\ldots, \varepsilon_{-1}, \varepsilon_0, \varepsilon_1, \ldots$ are independent (not necessarily identically distributed) and random variables $\epsilon_i, i \in \mathbb{Z}$ satisfy

$$\epsilon_i = g_i(\ldots, \epsilon_{i-1}, \varepsilon_{i-2}, \varepsilon_i)$$

(3)
In other words, \( \epsilon_i \) is measurable with the \( \sigma \)-field generated by \( \ldots, \epsilon_{i-2}, \epsilon_{i-1}, \epsilon_i \). The subscript \( i \) means that \( g_i \) can be different with respect to different \( i \). This system was first introduced by Wu [26]. Then it became an ubiquitous condition, see Wu and Wu [27], Zhang and Wu [28, 29, 22], Zhou [30, 31], Wang and Shao [32], Braumann et al. [16] and their reference.

For a random variable \( X \) and a number \( m \geq 1 \), denote \( \|X\|_m = (\mathbb{E}|X|^m)^{1/m} \). For any \( i \), define \( e_i \) as the random variable being independent with \( e_j, j \in \mathbb{Z} \), and having the same distribution as \( \epsilon_i \). Moreover, \( \ldots, e_{i-1}^\dagger, e_{i}^\dagger \) are mutually independent. Define the filter \( \mathcal{F}_i \) as the \( \sigma \)-field generated by \( \ldots, \epsilon_{i-1}, \epsilon_i \) and \( \mathcal{F}_{i,j} \) as the \( \sigma \)-field generated by \( \epsilon_{i-j}, \epsilon_{i-j+1}, \ldots, \epsilon_i \). Here \( i \in \mathbb{Z} \) and \( j \geq 0 \). Then \( \epsilon_i \) is \( \mathcal{F}_i \) measurable. Suppose a random variable \( X = f(\ldots, \epsilon_{i-1}, \epsilon_i) \) for a fixed \( i \), define

\[
X(j) = \begin{cases} 
  f(\ldots, \epsilon_{i-j-2}, \epsilon_{i-j-1}, e_{i-j}^\dagger, e_{i-j+1}, \ldots, \epsilon_i) & \text{if } j \geq 0 \\
  X & \text{if } j < 0
\end{cases} \tag{4}
\]

Define

\[
\delta_{i,j,m} = \|\epsilon_i - \epsilon_i(j)\|_m \text{ and } \delta_{i,m} = \sup_{j \in \mathbb{Z}} \delta_{i,j,m}
\tag{5}
\]

According to (1), \( \delta_{i,j,m} = \delta_{j,m} = 0 \) if \( j < 0 \).

This paper uses the notation \( \forall \) for ‘for all’, and \( \exists \) for ‘there exists’. The notations \( O(\cdot), o(\cdot), O_p(\cdot), o_p(\cdot) \) have the same meaning as in section 1.5.1, Shao [33], i.e., two numerical sequences \( \{a_n\}_{n=1}^{\infty}, \{b_n\}_{n=1}^{\infty} \) satisfy \( a_n = O(b_n) \) if \( \exists \) a constant \( C > 0 \) such that \( |a_n| \leq C|b_n| \) for all \( n \); and \( a_n = o(b_n) \) if \( |a_n|/|b_n| \to 0 \) as \( n \to \infty \). Two random variables sequences \( X_n, Y_n \) satisfy \( X_n = O_p(Y_n) \) if \( \forall \) given \( 0 < \varepsilon < 1, \exists \) a constant \( C > 0 \) such that \( \text{Prob}((|X_n| \leq C|Y_n|) \geq 1 - \varepsilon \) for any \( n \); and \( X_n = o_p(Y_n) \) if \( |X_n|/|Y_n| \to_p 0 \) as \( n \to \infty \). For two numbers \( x, y \), denote \( x \vee y = \max(x, y) \) and \( x \wedge y = \min(x, y) \). For a vector \( a = (a_1, \ldots, a_p)^T \in \mathbb{R}^p \) and \( k \geq 1 \), define the vector norm \( |a|_k = (\sum_{i=1}^{p} |a_i|^k)^{1/k} \). For a matrix \( A \), define the matrix 2-norm as \( |A|_2 = \sup_{|a|=1} |Aa|_2 \). Define \( \text{Prob}^*(\cdot) = \text{Prob}(\cdot|X_1, \ldots, X_n) \) and \( \mathbf{E}^* = \mathbf{E} \cdot |X_1, \ldots, X_n | \) as the probability and the expectation in the bootstrap world, i.e., conditional on the observed data.

This paper supposes the weakly stationary (section 1.3.2 in Brockwell and Davis [2]) random variable sequence \( X_i, i \in \mathbb{Z} \) satisfies an autoregressive representation

\[
X_i = \epsilon_i + \sum_{j=1}^{p} a_j X_{i-j}, \text{ here } P(x) = 1 - \sum_{j=1}^{p} a_j x^j \neq 0 \text{ for } x \in \mathbb{C}, |x| \leq 1 \tag{6}
\]
Here $p \geq 1$ is a constant and $a_j \in \mathbb{R}, j = 1, \ldots, p$ are fixed numbers. Besides, this paper applies the following assumptions:

**Assumptions**

1. $\epsilon_i, i \in \mathbb{Z}$ are white noise, i.e., $\mathbb{E}\epsilon_i = 0$; $\mathbb{E}\epsilon_i\epsilon_j = 0$ if $i \neq j$; $\mathbb{E}\epsilon_i^2 = \sigma^2$ for a constant $\sigma > 0$. Here $i, j \in \mathbb{Z}$.

2. $\exists$ constants $m \geq 8, \alpha > 1$ such that

$$\sup_{k=0,1,\ldots} (1 + k)^{\alpha} \sum_{j=k}^{\infty} \delta_{j,m} < \infty \quad \text{and} \quad \sup_{i \in \mathbb{Z}} \|\epsilon_i\|_m < \infty$$

(7)

3. $X_i, i \in \mathbb{Z}$ are weakly stationary, i.e., $\mathbb{E}X_i = 0$ and $\mathbb{E}X_iX_{i+j} = \mathbb{E}X_0X_j = \gamma_j \in \mathbb{R}$. Here $i \in \mathbb{Z}, j \geq 0$.

4. $X_i, i \in \mathbb{Z}$ and the coefficients $a_j, j = 1, 2, \ldots, p$ satisfy (6)

5. The matrix $\Gamma = \{\gamma_{|i-j|}, i,j=1,2,\ldots,p\}$ is non-singular. Besides, $\exists$ a constant $c_\zeta > 0$ such that the minimum eigenvalue of the matrix

$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \sum_{j=k+1}^{n} \mathbb{E}(X_{i1}X_{i1-k} - \gamma_k) \times (X_{i2}X_{i2-j} - \gamma_j) \right\}_{k,j=0,\ldots,p}$$

is greater than $c_\zeta$ for sufficiently large $n$.

The following lemma indicates that $X_i$ is causal, i.e., satisfies definition 3.1.3 in Brockwell and Davis [2].

**Lemma 1** (Theorem 3.1.1 in Brockwell and Davis [2])

Suppose assumption 1 to 4, then $\exists$ constants $\psi_j, j = 0, 1, 2, \ldots$ such that

$$X_i = \sum_{j=0}^{\infty} \psi_j \epsilon_{i-j} \text{ almost surely}$$

(8)

Here $\psi_j$ is determined by the relation $P(x) = \sum_{j=0}^{\infty} \psi_j x^j = 1$ for $|x| \leq 1$. In particular, $\exists \delta > 0$ and a constant $C$ such that $|\psi_j| \leq C(1 + \delta)^{-j}$.

A direct corollary of lemma 1 is

$$\sup_{i \in \mathbb{Z}} \|X_i\|_m \leq \sup_{i \in \mathbb{Z}} \sum_{j=0}^{\infty} |\psi_j| \times \|\epsilon_{i-j}\|_m \leq \sum_{j=0}^{\infty} |\psi_j| \times \sup_{i \in \mathbb{Z}} \|\epsilon_i\|_m < \infty$$

(9)

**Example 2** (An example satisfying (7))

Suppose $\epsilon_i, i \in \mathbb{Z}$ are independent random variables such that $\|\epsilon_i\|_m = 1$ and $a_{ij}, i, j \in \mathbb{Z}$ are real numbers such that $(1 + |j|^n + 1) \times |a_{ij}| < C < \infty$ for $\forall i, j \in \mathbb{Z}$. Define $X_i = \sum_{j=0}^{\infty} a_{ij} \epsilon_{i-j}$. From Kolmogorov’s three-series theorem, $X_i$ exists almost surely. Notice that $\|X_i - X_i(j)\|_m \leq 2|a_{ij}| < \frac{2C}{(1+j|a_{ij}|^{n+1})}$. So (7) is satisfied.
Remark 1

1. If we assume that $\epsilon_i$ are stationary, then $\delta_{j,m} = ||\epsilon_j - \epsilon_j(j)||_m$. In this case \(\delta\) coincides with (2.8) in Wu and Wu [27]. An alternative assumption for non-stationary time series is the 'locally stationary' assumption, see Zhang and Wu [22]. They required that the change in marginal distributions for consecutive data be asymptotically negligible. So the locally stationary assumption cannot handle the innovations like case 3 in example 1.

2. For any real numbers $a_j, j = 0, 1, ..., p$, from theorem 1 (in section 4) we have

$$\|\sum_{j=0}^{p} a_j \frac{1}{\sqrt{n}} \sum_{i=j+1}^{n} (X_i X_{i-j} - \gamma_j)\|_{m/2} = O\left(\left\{\sum_{j=0}^{p} a_j^2\right\}^{1/2}\right)$$

so the largest eigenvalue of the matrix $\{\frac{1}{n} \sum_{i=1}^{n} (X_{i-k} - \gamma_k) \times (X_{i-l} - \gamma_l)\}_{k,j=0,...,p}$ has order $O(1)$.

3. Some corollaries of assumption 2

Suppose random variables $\epsilon_i, i \in Z$ satisfy (3), (7) and $E\epsilon_i = 0$. For any coefficients $a_i, i = 1, 2, ..., n$ and $s \in Z$, define $\epsilon_{i,j} = E\epsilon_i|F_{i,j}$ and $M_{i,s} = \sum_{j=n+1-i}^{n} a_j (\epsilon_{j,s} - \epsilon_{j,s-1})$.

Then $M_{i,s}$ is measurable in the filter $F_{n,j+i-1}$ and $M_{j+1,s} - M_{i,s} = a_{n-1}(\epsilon_{n-i,s} - \epsilon_{n-i,s-1})$.

Apply $\pi - \lambda$ theorem to the $\lambda-$system

$$\{A \in F_{n,i+s-1} : E\epsilon_{n-i,s} \times 1_A = E\epsilon_{n-i,s-1} \times 1_A\}$$

and the $\pi-$system $\{A_n \times A_{n-1} \times ... \times A_{n-s-1}\}$. Here $A_i$ is generated by $\epsilon_i$. We have $E(\epsilon_{n-i,s} - \epsilon_{n-i,s-1})|F_{n,j+i-1} = 0$ almost surely. In particular, $\{M_{i,s}\}_{i=1,2,...,n}$ form a martingale. From Burkholder’s inequality (theorem 1.1 in [34]),

$$\|M_{n,s}\|_m \leq C \sqrt{\sum_{i=1}^{n} a_i^2 (\epsilon_{i,s} - \epsilon_{i,s-1})^2}_m/2 \leq C \sqrt{\sum_{i=1}^{n} a_i^2 \times \max_{i=1,...,n} \|\epsilon_{i,s} - \epsilon_{i,s-1}\|_m}$$

Here $C$ is a constant depending on $m$. Since

$$\|\epsilon_{i,s} - \epsilon_{i,s-1}\|_m = ||E(\epsilon_i - \epsilon_i(s))|F_{i,s}\|_m \leq \delta_{i,s,m}$$
Combine with theorem 2 in \cite{35},

\[
\| \sum_{i=1}^{n} a_i \epsilon_i \|_m \leq \| \sum_{i=1}^{n} a_i \epsilon_i \|_m + \sum_{s=1}^{\infty} \| M_{n,s} \|_m
\]

\[
\leq C \sqrt{\sum_{i=1}^{n} a_i^2 \| \epsilon_i \|_2^2} + C \sum_{i=1}^{n} a_i^2 \sum_{s=1}^{\infty} \delta_{s,m} = O \left( \sqrt{\sum_{i=1}^{n} a_i^2} \right)
\]

(14)

Here \( C \) is a constant. For a fixed integer \( k \geq 0 \), define \( \zeta_{i,k} = \epsilon_i \epsilon_i + k - E \epsilon_i \epsilon_i + k \). For fixed integers \( t, j, k \geq 0 \), define \( \nu_{i,t,j,k} = \zeta_{i,t,j} \zeta_{i} + t,k - E \zeta_{i,t,j} \zeta_{i} + t,k \).

\[
\| \epsilon_i \epsilon_i + k \|_m/2 \leq \| \epsilon_i \|_m \| \epsilon_i + k \|_m \Rightarrow \sup_{i \in \mathbb{Z}, k \geq 0} \| \epsilon_i \epsilon_i + k \|_m/2 < \infty
\]

(15)

In particular, this implies that \( \zeta_{i,k} \) and \( \nu_{i,t,j,k} \) are well-defined.

We have the following lemma:

**Lemma 2**

Suppose random variables \( \epsilon_i, i \in \mathbb{Z} \) satisfy \( \text{(3)} \) and \( \text{(7)} \).

1. For any fixed \( k \geq 0 \), define \( \Delta_{(i,k),j,m/2} = \| \zeta_{i,k} - \zeta_{i,k}(j) \|_m/2 \) and \( \Delta_{k,j,m/2} = \sup_{i \in \mathbb{Z}} \Delta_{(i,k),j,m/2} \). Then

\[
\sup_{j=0,1,...} (1+j)^{\alpha} \sum_{l=j}^{\infty} \Delta_{k,l,m/2} < \infty
\]

(16)

2. For any fixed \( k, j \geq 0 \), define \( \Theta_{(i,t,j,k),l,m/4} = \| \nu_{i,t,j,k} - \nu_{i,t,j,k}(l) \|_{m/4} \). Then

\[
\sup_{i \in \mathbb{Z}, t=0,1,...} \sum_{l=0}^{\infty} \Theta_{(i,t,j,k),l,m/4} < \infty
\]

(17)

If the innovations satisfy \( \text{(7)} \), then lemma \( \text{(3)} \) shows that the autoregressive process \( X_i, i \in \mathbb{Z} \) also satisfies \( \text{(7)} \).

**Lemma 3**
Suppose assumption 1 to 4. Define $\Delta_{i,j,m} = \|X_i - X_i(j)\|_m$, then we have
\[
\sup_{k=0,1,\ldots} \sum_{j=k}^{\infty} \sup_{i \in \mathbb{Z}} \Delta_{i,j,m} < \infty
\] (18)

Notably, from lemma 1
\[
\sup_{i \in \mathbb{Z}} \|X_i\|_m \leq \sum_{j=0}^{\infty} |\psi_j| \times \sup_{i \in \mathbb{Z}} \|\epsilon_i\|_m
\] (19)
so $\sup_{i \in \mathbb{Z}} \|\epsilon_i\|_m < \infty \Rightarrow \sup_{i \in \mathbb{Z}} \|X_i\|_m < \infty$.

**Remark 2**

A real-world time series $Z_i, i \in \mathbb{Z}$ may not have a linear expression with i.i.d. innovations.

If a statistician fits an AR(p) model on $Z_i$, the model is wrong. However, is it possible for this wrong model to fulfill the requirement of statistical inference, i.e., provide a consistent estimator for the autoregressive coefficients and make consistent confidence intervals?

Classical theories and methods rely on the innovations being i.i.d., so they fail to fulfill this requirement (as illustrated in example 1 and table 2).

For a fixed lag $p$, define $\hat{Z}_i = \sum_{j=1}^{p} a_j Z_{i-j}$. Here $\Gamma(a_1, \ldots, a_p)^T = (\gamma_1, \ldots, \gamma_p)^T$, $\Gamma$ is defined in assumption 5. Suppose $\Gamma$ is non-singular, then the solution $(a_1, \ldots, a_p)^T$ exists. Define $\delta_i = Z_{i} - \hat{Z}_i$ and set $a_0 = 1$. Suppose $Z_i$ satisfies assumption 2,
\[
\|\delta_i - \hat{\delta}_i(j)\|_m \leq \sum_{k=0}^{p} |a_k| \times \|Z_{i-k} - Z_{i-k}(j-k)\|_m
\]
\[
\Rightarrow \sum_{k=j}^{\infty} \sup_{i \in \mathbb{Z}} \|\delta_i - \hat{\delta}_i(k)\|_m \leq \sum_{l=0}^{p} |a_l| \times \sum_{k=0}^{\infty} \sup_{i \in \mathbb{Z}} \|Z_{i} - Z_{i}(k-l)\|_m
\] (20)
\[
\leq \sum_{l=0}^{p} \frac{C|a_l|}{(1 + 0 \vee (j-l))^{\alpha_l}}.
\]
Here $C$ is a constant. Therefore, as long as the lag $p$ is not very large, the innovation $\delta_i$ should still satisfy assumption 2.

On the other hand, suppose $i_2 < i_1$. Then $E\delta_{i_2} Z_{i_1} = 0$ for $i_2 = i_1 - 1, \ldots, i_1 - p$. Suppose
In other words, the Yule-Walker equation holds. So we define the statistics

\[ \delta_{i,j}^2 \leq \frac{\sum_{j=0}^{p} |a_j| \times |\mathbf{E}Z_{i,j}Z_{i,j}|}{\sum_{j=0}^{p} |a_j| \|Z_{i,j}\|_2} \leq \sum_{j=0}^{p} |a_j| \|Z_{i,j}\|_2 \leq \sum_{j=0}^{p} \frac{C|a_j|}{(1 + i_1 + i_2 - j)^{\alpha_c}} \]

(21)

So

\[ |\mathbf{E} \delta_{i,j} \hat{\delta}_{i,j}| \leq \sum_{k=0}^{p} |a_k| \times |\mathbf{E} \hat{\delta}_{i,j} \hat{\delta}_{i,j}| \leq \sum_{k=i_2+p+1-i_1}^{p} \sum_{j=0}^{\lfloor p/2 \rfloor} \frac{C|a_ja_k|}{(1 + i_1 + k - i_2 - j)^\alpha_c} \]

\[ \leq \frac{C}{(p/2 + 2)^\alpha_c} \sum_{k=i_2+p+1-i_1}^{p} \sum_{j=0}^{\lfloor p/2 \rfloor} |a_ja_k| + C \sum_{k=i_2+p+1-i_1}^{p} \sum_{j=0}^{\lfloor p/2 \rfloor} |a_ja_k| \]

(22)

Here \( |x| \) denotes the largest integer that is smaller than or equal to \( x \). If we impose additional assumptions on \( a_j \) (e.g., \( (1 + j)^{\alpha_c+1}|a_j| \leq C \) for any \( j \)), then \( \mathbf{E} \hat{\delta}_{i,j} \hat{\delta}_{i,j} \approx 0 \) and the innovations behave like a white noise series as \( p \to \infty \). According to (22), a statistician needs to fine-tune the lag \( p \) to maintain assumption 2 and make the innovations behave like a white noise series. But after doing that, our method can be applied to the ‘wrong model’ and generate a consistent confidence interval.

### 4 Consistency and Gaussian approximation

Define

\[ \gamma_i = \frac{1}{n} \sum_{j=i+1}^{n} X_jX_{j-i}, \quad \text{here we assume } n > p \text{ and } i = 0, 1, \ldots, p \]

(23)

Then define the estimator \( \hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2, \ldots, \hat{\gamma}_p)^T \) and the matrix \( \hat{\Gamma} = \{\hat{\gamma}_{i-j}\}_{i,j=1,\ldots,p} \). From assumption 1 to 5 and lemma \[ \mathbf{E}\epsilon_jX_j = \sum_{k=0}^{\infty} \psi_k \mathbf{E}\epsilon_j\epsilon_{j-k} = 0 \text{ for } j < i. \]

From assumption 1 to 5 and lemma \[ \mathbf{E}\epsilon_jX_j = \sum_{k=0}^{\infty} \psi_k \mathbf{E}\epsilon_j\epsilon_{j-k} = 0 \text{ for } j < i. \]

So

\[ \mathbf{E}\hat{\gamma}_i = \frac{n - i}{n} \gamma_i \text{ and } \gamma_j = \mathbf{E}X_jX_0 = \sum_{k=1}^{p} a_k \gamma_{j-k} \text{ for } j \geq 1 \]

(24)

In other words, the Yule-Walker equation holds. So we define the statistics

\[ \hat{a} = (\hat{a}_1, \ldots, \hat{a}_p)^T = \hat{\Gamma}^{-1}\hat{\gamma} \]

(25)
If \( \hat{\Gamma} \) is singular, then \( \hat{\Gamma}^{-1} \) stands for the pseudo-inverse of \( \hat{\Gamma} \). The first theorem involves demonstrating the consistency of \( \hat{\gamma} \).

**Theorem 1**

Suppose assumption 1 to 5. Then

\[
\|\hat{\gamma}_j - \gamma_j\|_{m/2} = O(1/\sqrt{n}) \quad \text{and} \quad |\hat{a}_k - a_k| = O_p(1/\sqrt{n})
\]

Here \( j = 0, 1, ..., p \) and \( k = 1, 2, ..., p \).

Then we focus on deriving a Gaussian approximation theorem for linear combinations of \( \hat{\gamma}_0, \ldots, \hat{\gamma}_p \). The Gaussian approximation theorems are powerful tools in high-dimensional statistics, see Chernozhukov et al. [36], Zhang and Politis [37] and Zhang and Wu [28]. Compared to the central limit theorem (e.g., theorem 1.15 in Shao [33]), the validity of a Gaussian approximation theorem only depends on the moments of random variables. Therefore, it is suitable for non-stationary time series as well.

For a given matrix \( B = \{b_{ij}\}_{i=1,...,p_1;j=0,1,...,p} \) such that \( \sum_{j=0}^{p_1} b_{ij}^2 > 0 \) and \( p_1 \leq p \), this paper focuses on constructing the simultaneous confidence interval for the parameter vector \((\sum_{j=0}^{p_1} b_{1j}\hat{\gamma}_j, \ldots, \sum_{j=0}^{p_1} b_{pj}\hat{\gamma}_j)^T\). So we need to derive its asymptotic distribution. Define the function \( H(x) \) as

\[
H(x) = \text{Prob} \left( \max_{i=1,2,\ldots,p_1} |\xi_i| \leq x \right), \quad x \in \mathbb{R}
\]

Here \((\xi_1, \ldots, \xi_{p_1})^T\) are joint normal random variables with mean \( E\xi_i = 0 \) and covariance

\[
E\xi_i\xi_j = \frac{1}{n} \sum_{j_1=0}^{p_1} \sum_{j_2=0}^{p_1} \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} b_{i1j_1}b_{i2j_2}E(X_{k_1}X_{k_1-j_1} - \hat{\gamma}_{j_1}) \times (X_{k_2}X_{k_2-j_2} - \hat{\gamma}_{j_2})
\]

For the time series \( X_i, i \in \mathbb{Z} \) is not assumed to be stationary, the matrix

\[
\left\{ \frac{1}{n} \sum_{j_1=0}^{p_1} \sum_{j_2=0}^{p_1} \sum_{k_1=0}^{n} \sum_{k_2=0}^{n} b_{i1j_1}b_{i2j_2}E(X_{k_1}X_{k_1-j_1} - \hat{\gamma}_{j_1}) \times (X_{k_2}X_{k_2-j_2} - \hat{\gamma}_{j_2}) \right\}_{k,j=0,\ldots,p}
\]

may not have a limit as \( n \to \infty \). Correspondingly, \( H \) changes when \( n \) varies. Yet theorem 2 shows that the difference between \( \max_{i=1,2,\ldots,p_1} |\sqrt{n}\sum_{j=0}^{p_1} b_{ij}(\hat{\gamma}_j - \gamma_j)|'s \) cumulative distribution function and \( H \) is asymptotically negligible.

**Theorem 2**

Suppose assumption 1 to 5. Then

\[
\sup_{x \in \mathbb{R}} |\text{Prob} \left( \max_{i=1,2,\ldots,p_1} |\sqrt{n}\sum_{j=0}^{p_1} b_{ij}(\hat{\gamma}_j - \gamma_j)| \leq x \right) - H(x) | = o(1)
\]
Here $B = \{b_{ij}\}_{i=1, \ldots, p_1, j=0, \ldots, p}$ is a fixed matrix such that $\sum_{j=0}^{p} b_{ij}^2 > 0$ for $i = 1, 2, \ldots, p_1$ and $p_1 \leq p$.

From theorem 1 and (5.8.4) in Horn and Johnson [38],

$$|(\bar{\eta} - \eta) - (\bar{\Gamma} - \Gamma)\eta - \Gamma^{-1} (\bar{\gamma} - \gamma)|_2 \leq |\bar{\Gamma} - \Gamma|_2 \times |\bar{\gamma} - \gamma|_2 = O_p(1/n) \quad (30)$$

From corollary 5.6.16 in [38], if $|\Gamma^{-1}(\bar{\Gamma} - \Gamma)|_2 < 1/2$, then

$$\bar{\Gamma}^{-1} = \left( \sum_{k=0}^{\infty} (-1)^k (\Gamma^{-1} (\bar{\Gamma} - \Gamma))^k \right) \Gamma^{-1} \quad (31)$$

Define the $p \times p$ matrix $T_i = \{t_{i,j,k}\}_{j,k=1, 2, \ldots, p}$ such that $t_{i,j,k} = 1$ if $j = k$ and 0 otherwise. Then define $\delta_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)^T$. For a fixed matrix $B = \{b_{ij}\}_{i=1, \ldots, p_1, j=1, \ldots, p}$, set $b_i = (b_{i1}, \ldots, b_{ip})^T$,

$$\sqrt{n}(B \bar{\eta} - B \eta) = \sqrt{n} \sum_{i=1}^{p} (\bar{\gamma}_i - \gamma_i) B \Gamma^{-1} \delta_i - \sqrt{n} \sum_{i=p+1}^{p-1} (\bar{\gamma}_i - \gamma_i) B \Gamma^{-1} T_i \Gamma^{-1} \gamma + O_p(1/\sqrt{n})$$

\[= \sqrt{n} \sum_{j=1}^{p} b_j (\bar{a}_j - a_j) = b_i^T \Gamma^{-1} \delta_i \times \sqrt{n}(\gamma_p - \gamma_0) - b_i^T \Gamma^{-1} T_0 \Gamma^{-1} \gamma \times \sqrt{n}(\gamma_0 - \gamma_0) \]

\[+ \sum_{j=1}^{p-1} b_j^T \Gamma^{-1} \delta_j - b_j^T \Gamma^{-1} (T_j + T_{-j}) \Gamma^{-1} \gamma \times \sqrt{n}(\gamma_0 - \gamma_0) + O_p(1/\sqrt{n}) \quad (32)\]

Therefore we have the following corollary.

**Corollary 1**

Suppose assumption 1 to 5. Then

$$\sup_{x \in \mathbb{R}} |\text{Prob} \left( \max_{i=1, \ldots, p_1} \left| \sqrt{n} \sum_{j=1}^{p} b_j (\bar{a}_j - a_j) \right| \leq x \right) - \mathcal{H}(x) | = o(1) \quad (33)$$

Here $B = \{b_{ij}\}_{i=1, \ldots, p_1, j=1, \ldots, p}$ is a fixed matrix such that

$$(b_i^T \Gamma^{-1} \delta_p)^2 + (b_i^T \Gamma^{-2} \gamma)^2 + \sum_{j=1}^{p-1} (b_j^T \Gamma^{-1} \delta_j - b_j^T \Gamma^{-1} (T_j + T_{-j}) \Gamma^{-1} \gamma)^2 > 0 \quad (34)$$

for $i = 1, 2, \ldots, p_1$, $p_1 \leq p$, and $\mathcal{H}(x) = \text{Prob}(\max_{i=1, \ldots, p_1} |\xi_i| \leq x)$. $\xi_1, \ldots, \xi_{p_1}$ are joint
normal random variables with $E\xi_i = 0$,

$$E\xi_{i1}\xi_{i2} = \frac{1}{n} \sum_{j_1=0}^{p} \sum_{j_2=0}^{p} \sum_{k_1=1}^{n} \sum_{k_2=1}^{n} c_{i1,j_1} c_{i2,j_2} E(X_{k_1} X_{k_1-j_1} - \gamma_{j_1}) \times (X_{k_2} X_{k_2-j_2} - \gamma_{j_2})$$

(35)

$$c_{i0} = -b_i^T \Gamma^{-2} \gamma, \ c_{ip} = b_i^T \Gamma^{-1} \delta_p \ and \ c_{ij} = (b_i^T \Gamma^{-1} \delta_j - b_i^T \Gamma^{-1} (T_j + T_{-j}) \Gamma^{-1} \gamma) \ for \ j = 1, 2, ..., p - 1.$$

5 Bootstrap confidence interval & hypothesis testing

So far, we have derived the Gaussian approximation theorem for the Yule-Walker estimator. Then statisticians may try to estimate the Yule-Walker estimator’s variance and generate a consistent confidence interval based on the normal distribution.

Another idea is to use a computer intensive method, like subsampling or bootstrap [39], that generates a consistent confidence interval / performs hypothesis testing through simulations. The advantage of this idea is that it implicitly derives the estimator’s variance, so statisticians do not need to calculate anything. This section adopts the second idea and introduces a bootstrap algorithm, called the second order wild bootstrap. We are interested in testing the statistical hypothesis

$$null: Ba = c \ versus \ alternative \ Ba \neq c \quad (36)$$

Here $B$ is a known $p_1 \times p$ matrix and $c$ is a known $p_1 \times 1$ vector. Apart from assumption 1 to 5, this section requires an additional assumption.

Additional assumption

6. Suppose a function $K(\cdot) : \mathbb{R} \rightarrow [0, \infty)$ is symmetric and continuously differentiable. $K(0) = 1$, $\int_{\mathbb{R}} K(x) dx < \infty$. $K(x)$ is decreasing on $[0, \infty)$. Define the Fourier transformation of $K$ as $\mathcal{F}K(x) = \int_{\mathbb{R}} K(t) \exp(-2i \times \pi tx) dt$. Assume $\mathcal{F}K(x) \geq 0$ for $\forall x \in \mathbb{R}$, and $\int_{\mathbb{R}} \mathcal{F}K(x) dx < \infty$. Suppose a bandwidth parameter $k_n > 0$ satisfies $\lim_{n \to \infty} k_n = \infty$ and $k_n = o(\sqrt{n})$.

Remark 3

According to Shao [40] and the Fourier inversion theorem (Theorem 8.26 in Folland [41]).
\( \forall x = (x_1, ..., x_n)^T \in \mathbb{R}^n, \)
\[
\sum_{s=1}^{n} \sum_{j=1}^{n} x_s x_j K \left( \frac{s-j}{k_n} \right) = \int_{\mathbb{R}} \sum_{s=1}^{n} \sum_{j=1}^{n} x_s x_j F_K(z) \exp \left( 2\pi i z \frac{s-j}{k_n} \right) dz
\]
\[
= \int_{\mathbb{R}} F_K(z) \times \sum_{s=1}^{n} x_s \exp \left( 2\pi i z \frac{s}{k_n} \right) |z|^2 \geq 0
\]

so the matrix \( \{ K \left( \frac{s-j}{k_n} \right) \}_{s,j=1,2,...,n} \) is symmetric positive semi-definite. One satisfactory kernel \( K \) is \( K(x) = \exp(-x^2/2) \) whose Fourier transformation is \( \tilde{F}_K(z) = \sqrt{2\pi} \exp(-\pi^2 z^2) \).

Algorithm 1 presents the second order wild bootstrap.

**Algorithm 1:** second order wild bootstrap

**Input:** observations \( X_1, ..., X_n \), a kernel function \( K(\cdot) \), a bandwidth \( k_n > 0 \), the nominal coverage probability \( 1 - \alpha \), the number of bootstrap replicates \( T \), and the linear combination matrix \( B = \{ b_{ij} \}_{i=1,...,p \atop j=1,...,n} \).

**Additional input for testing (36):** The expected vector \( c = (c_1, ..., c_p)^T \)

1. Estimate coefficients \( \hat{\gamma}_0, ..., \hat{\gamma}_p \) and \( \hat{a}_1, ..., \hat{a}_p \) according to (24) and (25).
2. Generate joint normal distributed random variables \( \varepsilon_1, ..., \varepsilon_n \) such that \( \mathbb{E}^* \varepsilon_i = 0, \mathbb{E}^* \varepsilon_i \varepsilon_j = K \left( \frac{i-j}{k_n} \right) \).
3. Define
\[
\hat{\gamma}_i^* = \gamma_i + \frac{1}{n} \sum_{j=i+1}^{n} (X_j X_{j-i} - \gamma_i) \times \varepsilon_j, i = 0, 1, ..., p \quad (38)
\]

Then define \( \hat{\gamma}^* = (\hat{\gamma}_1^*, ..., \hat{\gamma}_p^*)^T \) and \( \hat{\gamma}^* = (\hat{\gamma}_1^*, ..., \hat{\gamma}_p^*)^T \). Calculate
\[
\hat{\gamma}_i^* = \frac{1}{n} \sum_{j=i+1}^{n} (X_j X_{j-i} - \gamma_i) \times \varepsilon_j, i = 0, 1, ..., p \quad (39)
\]

If \( \hat{\gamma}^* \) is singular, then \( \hat{\gamma}^*^{-1} \) represents the pseudo-inverse of \( \hat{\gamma}^* \).
4. Calculate the statistics \( \delta_b = \max_{i=1,...,p_1} \sqrt{n} | \sum_{j=1}^{p_1} b_{ij} \hat{a}_j - \sum_{j=1}^{p_1} b_{ij} \hat{a}_j | \)
5. Repeat step 2 to 4 for \( b = 1, 2, ..., T \). Then calculate the \( 1 - \alpha \) sample quantile \( C_{1-\alpha}^* \) of \( \delta_1, ..., \delta_B \).
6.a (for making confidence interval) The \( 1 - \alpha \) confidence interval for \( Bab \) will be
\[
\left\{ x \in \mathbb{R}^{p_1} : \max_{i=1,...,p_1} \left| x_i - \sum_{j=1}^{p} b_{ij} \hat{a}_j \right| \leq \frac{C_{1-\alpha}^*}{\sqrt{n}} \right\}
\]

6.b (for hypothesis testing) Reject the null hypothesis if
\[
\max_{i=1,...,p_1} \left| \sum_{j=1}^{p} b_{ij} \hat{a}_j - c_i \right| > \frac{C_{1-\alpha}^*}{\sqrt{n}}
\]

**Remark 4**
Algorithm 1 adopts the idea of the dependent wild bootstrap introduced by Shao [40]. However, it treats the sequence \{X_iX_{i-k} - \gamma_k\}_{i=k+1, \ldots, n} (k is a fixed number) as a new time series and resamples on this sequence rather than \(X_i\). After applying this modification, algorithm 1 is able to capture the second order structure of the sequence \{X_iX_{i-k} - \gamma_k\}_{i=k+1, \ldots, n}, which is the high order structure of the original time series.

According to theorem 1.2.1 in Politis et al. [39], the consistency of algorithm 1 is ensured if one can show that the distribution of the bootstrapped estimator \(\max_{i=1, \ldots, p}\sqrt{n} \left| \sum_{j=1}^{p} b_{ij} \tilde{a}_{ij} - \sum_{j=1}^{p} b_{ij} \tilde{a}_{ij} \right|\) converges to \(\mathcal{H}\) (defined in corollary 1) in probability. Theorem 3 justifies this result.

**Theorem 3**
Suppose assumption 1 to 6, then

\[
\sup_{x \in \mathbb{R}} |\text{Prob}^* \left( \max_{i=1, \ldots, p} \sqrt{n} \left| \sum_{j=1}^{p} b_{ij} \tilde{a}_{ij} - \sum_{j=1}^{p} b_{ij} \tilde{a}_{ij} \right| \leq x \right) - \mathcal{H}(x) | = o_p(1)
\]  

(42)

Here \(\mathcal{H}\) is defined in corollary 1 and \(\tilde{a}^* = (\tilde{a}_1^*, \ldots, \tilde{a}_p^*)^T\) is defined in algorithm 1.

6 Numerical result

Suppose \(e_i, i \in \mathbb{Z}\) are i.i.d. normal random variables with mean 0 and variance 1. Following example 1, we consider three types of white noise innovations: normal: \(\epsilon_i = e_i\), product normal: \(\epsilon_i = e_i e_{i-1}\) and non-stationary: \(\epsilon_{2i} = e_i, \epsilon_{2i-1} = e_i e_{i-1}\). Then we consider the following 4 time series:

1. AR(1) model \(X_i = 0.7X_{i-1} + \epsilon_i\)
2. AR(2) model \(X_i = 0.3X_{i-1} + 0.5X_{i-2} + \epsilon_i\)
3. MA(1) model \(X_i = \epsilon_i + 2.0\epsilon_{i-1}\)
4. non-linear model \(X_i = 0.6 \sin(X_{i-1}) + \epsilon_i\)

3 is a linear time series that is not invertible according to Keriss et al. [1], and 4 is considered in Fragkeskou and Paparoditis [25]. For 3 and 4 do not have explicit AR coefficients, we calculate them through simulating 50000 samples. The order of \(X_i\) in this situation is determined by the aic criterion implemented in R’s ‘ar’ function.

**Selecting bandwidth \(k_n\)**

Politis and White [42] introduced an automatic bandwidth selection algorithm and Shao
applied this algorithm to select the bandwidth of the dependent wild bootstrap. The R package ‘up’ implemented this method. Other methods include Fragkeskou and Paparoditis (also introduced in [25]). For convenience, we apply the Politis and White’s method. However, their method was decided for a stationary time series. So it may result in selecting a sub-optimal bandwidth.

Simulation results

We demonstrate the simulation results in table 2. To accelerate calculation, we apply the warp-speed algorithm introduced by Giacomini et al. to calculate the coverage probability. We compare the performance of the second order wild bootstrap and the AR bootstrap in table 2.

Despite different types of white noise innovations are incorporated to the time series model, the AR bootstrap’s 90% quantile remains the same for each model. Therefore, its confidence interval has correct coverage probability when the innovations are indeed i.i.d. but is too narrow when the innovations become dependent. In other words, the AR bootstrap fails to capture the high order structures of a time series that affect the width of a confidence interval.

On the other hand, the second order wild bootstrap’s 90% quantile is close to the AR bootstrap’s 90% quantile when the innovations are i.i.d., but is significantly larger than that when the innovations are dependent. Therefore, the impact of the time series’s high order structures on the confidence interval can be captured by the second order wild bootstrap.

7 Conclusion

Theoretical results and practical tools for fitting an AR($p$) model with i.i.d. innovations are well-established. However, when the innovations in the AR($p$) model is no longer i.i.d., the existing results fail. This paper establishes the consistency and the Gaussian approximation theorem for the Yule-Walker estimator when the innovations in the AR($p$) model is dependent and non-stationary. Moreover, it derives the second order wild bootstrap that automatically generates a consistent confidence interval for the Yule-Walker estimator.

When the innovations are dependent, the Yule-Walker estimator’s confidence interval will be affected by the high order structures of the time series. Therefore, classical methods like AR bootstrap fails to generate a consistent confidence interval. Since a general weakly
Table 2: Performance of bootstrap algorithms on various time series. This simulation chooses $B = I_p$, the identity matrix with dimension $p$ (equals lag) ‘AR’ abbreviates ‘AR bootstrap’ and ‘Wild’ abbreviates ‘the second order wild bootstrap’. ‘P’ represents the coverage probability derived by the warp speed algorithm and ‘C’ denotes the quantile of $\max_{i=1,...,p} \sqrt{n} |\hat{a}_i - \tilde{a}_i|$ (see algorithm [1]) calculated by the corresponding bootstrap algorithm. ‘Lag’ denotes the lag of the autoregressive model. For model 3 and 4, the lag is chosen by the aic criterion implemented in the R’s ‘ar’ function. The bandwidth $k_n$ for the second order wild bootstrap is selected based on Politis and White [42]. The nominal coverage probability is 90%.

| Model | Innovation | Sample size | Lag | Wild P | Wild C | AR P | AR C |
|-------|------------|-------------|-----|--------|--------|------|------|
| 1     | normal     | 500         | 1   | 81.9%  | 1.022  | 82.4%| 1.182|
|       | normal     | 1000        | 1   | 89.0%  | 1.209  | 90.6%| 1.296|
|       | product normal | 500    | 1   | 95.3%  | 1.564  | 73.9%| 1.160|
|       | product normal | 1000  | 1   | 91.7%  | 1.949  | 70.9%| 1.250|
|       | non-stationary | 500   | 1   | 88.1%  | 1.656  | 68.6%| 1.155|
|       | non-stationary | 1000  | 1   | 88.0%  | 1.512  | 65.4%| 1.139|
| 2     | normal     | 500         | 2   | 90.7%  | 1.734  | 87.6%| 1.671|
|       | normal     | 1000        | 2   | 92.0%  | 2.305  | 87.1%| 1.848|
|       | product normal | 500    | 2   | 92.2%  | 3.173  | 65.2%| 1.827|
|       | product normal | 1000  | 2   | 96.0%  | 2.494  | 73.0%| 1.722|
|       | non-stationary | 500   | 2   | 89.1%  | 3.248  | 62.5%| 1.604|
|       | non-stationary | 1000  | 2   | 88.7%  | 2.965  | 60.1%| 1.637|
| 3     | normal     | 500         | 4   | 98.7%  | 2.814  | 91.3%| 2.291|
|       | normal     | 1000        | 4   | 97.7%  | 2.746  | 91.5%| 2.344|
|       | product normal | 500    | 3   | 95.9%  | 4.111  | 73.9%| 2.196|
|       | product normal | 1000  | 7   | 98.6%  | 4.033  | 76.5%| 2.687|
|       | non-stationary | 500   | 3   | 93.9%  | 5.100  | 61.7%| 2.402|
|       | non-stationary | 1000  | 4   | 94.1%  | 4.438  | 55.1%| 2.356|
| 4     | normal     | 500         | 1   | 87.4%  | 1.262  | 87.2%| 1.379|
|       | normal     | 1000        | 1   | 89.3%  | 1.448  | 90.6%| 1.388|
|       | product normal | 500    | 1   | 92.3%  | 2.372  | 75.1%| 1.334|
|       | product normal | 1000  | 7   | 99.3%  | 4.523  | 79.3%| 2.667|
|       | non-stationary | 500   | 1   | 85.1%  | 2.011  | 64.5%| 1.303|
|       | non-stationary | 1000  | 1   | 85.2%  | 1.810  | 65.7%| 1.321|
stationary time series is likely to have a linear expression (because of Kreiss et al. and the Wold decomposition) with dependent innovations, our result can be a good choice to solve real-life problems.

References

[1] Jens-Peter Kreiss, Efstathios Paparoditis, and Dimitris N. Politis. On the range of validity of the autoregressive sieve bootstrap. *The Annals of Statistics*, 39(4):2103 – 2130, 2011.

[2] Peter J Brockwell and Richard A Davis. *Time Series: Theory and Methods*. Springer-Verlag, Berlin, Heidelberg, 1986.

[3] Y. Nardi and A. Rinaldo. Autoregressive process modeling via the lasso procedure. *Journal of Multivariate Analysis*, 102(3):528 – 549, 2011.

[4] Li Pan and Dimitris N. Politis. Bootstrap prediction intervals for linear, nonlinear and nonparametric autoregressions. *Journal of Statistical Planning and Inference*, 177:1 – 27, 2016.

[5] Fang Han, Huanran Lu, and Han Liu. A direct estimation of high dimensional stationary vector autoregressions. *Journal of Machine Learning Research*, 16(97):3115–3150, 2015.

[6] Sumanta Basu and George Michailidis. Regularized estimation in sparse high-dimensional time series models. *The Annals of Statistics*, 43(4):1535 – 1567, 2015.

[7] Jonas Krampe and Efstathios Paparoditis. Sparsity concepts and estimation procedures for high-dimensional vector autoregressive models. *Journal of Time Series Analysis*, 42(5-6):554–579, 2021.

[8] Jianqing Fan and Qiwei Yao. *Nonlinear Time Series*. Springer-Verlag, New York, 2003.

[9] Claudia Kirch and Joseph Tadjuidje Kamgaing. Testing for parameter stability in nonlinear autoregressive models. *Journal of Time Series Analysis*, 33(3):365–385, 2012.

[10] Zongwu Cai, Jianqing Fan, and Qiwei Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95(451):941–956, 2000.

[11] Rong Chen and Ruey S. Tsay. Functional-coefficient autoregressive models. *Journal of the American Statistical Association*, 88(421):298–308, 1993.
[12] Yingcun Xia and W. K. Li. On the estimation and testing of functional-coefficient linear models. *Statistica Sinica*, 9(3):735–757, 1999.

[13] Jens-Peter Kreiss. Bootstrap procedures for ar (\(\infty\)) — processes. In Karl-Heinz Jöckel, Günter Rothe, and Wolfgang Sendler, editors, *Bootstrapping and Related Techniques*, pages 107–113, Berlin, Heidelberg, 1992. Springer Berlin Heidelberg.

[14] Peter Bühlmann. Sieve bootstrap for time series. *Bernoulli*, 3(2):123 – 148, 1997.

[15] Efstatios Paparoditis. Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes. *Journal of Multivariate Analysis*, 57(2):277–296, 1996.

[16] Alexander Braumann, Jens-Peter Kreiss, and Marco Meyer. Simultaneous inference for autocovariances based on autoregressive sieve bootstrap. *Journal of Time Series Analysis*, 42(5-6):534–553, 2021.

[17] Zeng Li, Clifford Lam, Jianfeng Yao, and Qiwei Yao. On testing for high-dimensional white noise. *The Annals of Statistics*, 47(6):3382 – 3412, 2019.

[18] Pramita Bagchi, Vaidotas Characiejus, and Holger Dette. A simple test for white noise in functional time series. *Journal of Time Series Analysis*, 39(1):54–74, 2018.

[19] Han Xiao and Wei Biao Wu. Portmanteau test and simultaneous inference for serial covariances. *Statistica Sinica*, 24(2):577–599, 2014.

[20] Holger Dette, Weichi Wu, and Zhou Zhou. Change point analysis of second order characteristics in non-stationary time series. (arXiv: 1503.08610), 2015.

[21] Jiang Wang and Dimitris N. Politis. Consistent autoregressive spectral estimates: Non-linear time series and large autocovariance matrices. *Journal of Time Series Analysis*, 42(5-6):580–596, 2021.

[22] Danna Zhang and Wei Biao Wu. Convergence of covariance and spectral density estimates for high-dimensional locally stationary processes. *The Annals of Statistics*, 49(1):233 – 254, 2021.

[23] Timothy L. McMurry and Dimitris N. Politis. Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. *Journal of Time Series Analysis*, 31(6):471–482, 2010.
[24] Carsten Jentsch and Dimitris N. Politis. Covariance matrix estimation and linear process bootstrap for multivariate time series of possibly increasing dimension. *The Annals of Statistics*, 43(3):1117 – 1140, 2015.

[25] Maria Fragkeskou and Efstathios Paparoditis. Extending the range of validity of the autoregressive (sieve) bootstrap. *Journal of Time Series Analysis*, 39(3):356–379, 2018.

[26] Wei Biao Wu. Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences*, 102(40):14150–14154, 2005.

[27] Wei-Biao Wu and Ying Nian Wu. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electron. J. Statist.*, 10(1):352–379, 2016.

[28] Danna Zhang and Wei Biao Wu. Gaussian approximation for high dimensional time series. *The Annals of Statistics*, 45(5):1895 – 1919, 2017.

[29] Danna Zhang and Wei Biao Wu. Asymptotic theory for estimators of high-order statistics of stationary processes. *IEEE Transactions on Information Theory*, 64(7):4907–4922, 2018.

[30] Zhou Zhou. Inference of weighted $V$-statistics for nonstationary time series and its applications. *The Annals of Statistics*, 42(1):87 – 114, 2014.

[31] Zhou Zhou. Inference for non-stationary time series regression with or without inequality constraints. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(2):349–371, 2015.

[32] Runmin Wang and Xiaofeng Shao. Hypothesis testing for high-dimensional time series via self-normalization. *The Annals of Statistics*, 48(5):2728 – 2758, 2020.

[33] Jun Shao. *Mathematical Statistics*. Springer-Verlag New York, 2003.

[34] D. L. Burkholder, B. J. Davis, and R. F. Gundy. Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Probability Theory*, pages 223–240, Berkeley, Calif., 1972. University of California Press.

[35] Peter Whittle. Bounds for the moments of linear and quadratic forms in independent variables. *Theory of Probability & Its Applications*, 5(3):302–305, 1960.
[36] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *Ann. Statist.*, 41(6):2786–2819, 12 2013.

[37] Yunyi Zhang and D. N. Politis. Ridge regression revisited: Debiasing, thresholding and bootstrap. (arXiv:2009.08071), 2020.

[38] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, 2012.

[39] Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer-Verlag New York, 1999.

[40] Xiaofeng Shao. The dependent wild bootstrap. *Journal of the American Statistical Association*, 105(489):218–235, 2010.

[41] Gerald B. Folland. *Real Analysis*. John Wiley & Sons, 2nd edition, 1999.

[42] Dimitris N. Politis and Halbert White. Automatic block-length selection for the dependent bootstrap. *Econometric Reviews*, 23(1):53–70, 2004.

[43] Tristen Hayfield and Jeffrey S. Racine. Nonparametric econometrics: The np package. *Journal of Statistical Software*, 27(5), 2008.

[44] Maria Fragkeskou and Efstathios Paparoditis. Inference for the fourth-order innovation cumulant in linear time series. *Journal of Time Series Analysis*, 37(2):240–266, 2016.

[45] Raffaella Giacomini, Dimitris N. Politis, and Halbert White. A warp-speed method for conducting monte carlo experiments involving bootstrap estimators. *Econometric Theory*, 29(3):567 – 589, 2013.

[46] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Gaussian approximations and multiplier bootstrap for maxima of sums of high-dimensional random vectors. *The Annals of Statistics*, 41(6):2786 – 2819, 2013.

[47] Victor Chernozhukov, Denis Chetverikov, and Kengo Kato. Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields*, 162:47–70, 2015.
A Preliminary

This section introduces some preliminary definitions and results used in this paper.

For any $\tau, \psi > 0, z \in \mathbb{R}$, define $F_{\tau}(x_1, ..., x_n) = \frac{1}{\tau} \log \left( \sum_{i=1}^{n} \exp(\tau x_i) \right)$; $G_{\tau}(x_1, ..., x_n) = \frac{1}{\tau} \log \left( \sum_{i=1}^{n} \exp(\tau x_i) + \sum_{i=1}^{n} \exp(-\tau x_i) \right) = F_{\tau}(x_1, ..., x_n) - F_{\tau}(0, ..., 0)$; $g_{\psi}(x) = (1-\min(1, \max(x, 0))^4$; $g_{\psi,z}(x) = g_{\psi}(\psi(x-z))$; and $h_{\tau,\psi,z}(x_1, ..., x_n) = g_{\psi,z}(G_{\tau}(x_1, ..., x_n))$. From lemma A.2 and (8) in Chernozhukov et al. [36] and (S1) to (S5) in Zhang and Wu [28], $g_{\psi} = \sup_{x \in \mathbb{R}} |g_{\psi}(x)| + |g_{\psi'}(x)| < \infty$. And $1_{x \leq \tau} \leq g_{\psi,z}(x) \leq 1_{x \leq \tau + 1/\psi}$. Also $\sup_{z \in \mathbb{R}} |g_{\psi,z}(x)| \leq g_{\psi,\psi} \sup_{x \in \mathbb{R}} |g_{\psi,z}(x)| \leq g_{\psi}^2$ and $\sup_{z \in \mathbb{R}} |g_{\psi,z}(x)| \leq g_{\psi}^3$.

Define
\[
\begin{aligned}
\partial_{i} &= \frac{\partial}{\partial x_i}, \\
\partial_{i,j} &= \frac{\partial^2}{\partial x_i \partial x_j} \quad \text{and} \\
\partial_{i,j,k} &= \frac{\partial^3}{\partial x_i \partial x_j \partial x_k}
\end{aligned}
\]
we have $\partial_{i} F_{\tau} \geq 0, \sum_{i=1}^{n} \partial_{i} F_{\tau} = 1, \sum_{i=1}^{n} \sum_{j=1}^{n} |\partial_{i,j} F_{\tau}| \leq 2\tau$ and $\sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} |\partial_{i,j,k} F_{\tau}| \leq 6\tau^2$. Besides, $F_{\tau}(x_1, ..., x_n) - \frac{\log(n)}{\tau} \leq \max_{i=1, ..., n} |x_i| \leq F_{\tau}(x_1, ..., x_n)$. Therefore,
\[
G_{\tau}(x_1, ..., x_n) - \frac{\log(2n)}{\tau} \leq \max_{i=1, ..., n} |x_i| \leq G_{\tau}(x_1, ..., x_n)
\]
and
\[
\partial_{i} G_{\tau} = \partial_{i} F_{\tau}(x_1, ..., x_n, -x_1, ..., -x_n) - \partial_{i+n} F_{\tau}(x_1, ..., x_n, -x_1, ..., -x_n) \Rightarrow \sum_{i=1}^{n} |\partial_{i} G_{\tau}| \leq 1
\]
Similarly $\partial_{i,j} G_{\tau} = \partial_{i,j} F_{\tau} - \partial_{i+n,j} F_{\tau} - \partial_{i,j+n} F_{\tau} + \partial_{i+n,j+n} F_{\tau} \Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} |\partial_{i,j} G_{\tau}| \leq 2\tau$.

and
\[
\partial_{i,j,k} G_{\tau} = \partial_{i,j,k} F_{\tau} + \partial_{i+n,j+k} F_{\tau} + \partial_{i+n,j+n+k} F_{\tau} + \partial_{i+n,j+n+k} F_{\tau} - \partial_{i+n,j+k} F_{\tau} - \partial_{i,j+n+k} F_{\tau} - \partial_{i+n,j+n+k} F_{\tau} \Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} |\partial_{i,j,k} G_{\tau}| \leq 6\tau^2
\]

Since $\partial_{i} h_{\tau,\psi,z} = g_{\psi,z}(G_{\tau}(x_1, ..., x_n)) \partial_{i} g_{\psi}$, we get $\sum_{i=1}^{n} |\partial_{i} h_{\tau,\psi,z}| \leq g_{\psi} \psi$.

\[
\begin{aligned}
\partial_{i,j} h_{\tau,\psi,z} &= g_{\psi,z}(G_{\tau}(x_1, ..., x_n)) \partial_{i,j} G_{\tau} + g_{\psi,z}(G_{\tau}(x_1, ..., x_n)) \partial_{i,j} G_{\tau} \\
&\Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} |\partial_{i,j} h_{\tau,\psi,z}| \leq g_{\psi} \psi^2 + 2g_{\psi} \psi \tau
\end{aligned}
\]

23
and
\[
\partial_{i,j,k}h_{r,\psi,z} = g_{r,z}''(G_r(x_1, \ldots, x_n))\partial_i G_r \partial_j G_r \partial_k G_r + g_{r,z}'(G_r(x_1, \ldots, x_n))\partial_{i,j,k} G_r
\]
\[+ g_{r,z}'(G_r(x_1, \ldots, x_n)) (\partial_{i,j} G_r \partial_k G_r + \partial_{j,k} G_r \partial_i G_r + \partial_{i,k} G_r \partial_j G_r)\]
(48)

\[\Rightarrow \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{k=1}^{n} |\partial_{i,j,k} h_{r,\psi,z}| \leq g_r \psi^3 + 6g_r \psi^2 + 6g_r \psi^2 \tau\]

We end this section by introducing a lemma about joint normal random variables from Chernozhukov et al. [36].

**Lemma 4**

Suppose \( \epsilon = (\epsilon_1, \ldots, \epsilon_n)^T \) are joint normal random variables, \( E\epsilon = 0, E\epsilon \epsilon^T = \Sigma = (\sigma_{ij})_{i,j=1,\ldots,n} \);
and \( \exists 0 < c_0 \leq C_0 < \infty \) such that \( c_0 \leq \sigma_{ii} \leq C_0, \ i = 1, \ldots, n \). Then

\[
\sup_{x \in \mathbb{R}} |\text{Prob} \left( \max_{i=1,\ldots,n} |\epsilon_i| \leq x + \delta \right) - \text{Prob} \left( \max_{i=1,\ldots,n} |\epsilon_i| \leq x \right) | 
\leq C_0 \left( 1 + \sqrt{\log(n)} + \sqrt{\log(\delta)} \right)
\]
(49)

In addition suppose \( \xi = (\xi_1, \ldots, \xi_n)^T \) are joint normal random variables with \( E\xi = 0, E\xi \xi^T = \Sigma^* = (\sigma_{ij}^*)_{i,j=1,\ldots,n} \), and \( \Delta = \max_{i,j=1,\ldots,n} |\sigma_{ij} - \sigma_{ij}^*| < 1 \). Then

\[
\sup_{x \in \mathbb{R}} |\text{Prob} \left( \max_{i=1,\ldots,n} |\xi_i| \leq x \right) - \text{Prob} \left( \max_{i=1,\ldots,n} |\xi_i| \leq x \right) | 
\leq C^* \left( \Delta^{1/3} \times (1 + \log^3(n)) + \frac{\Delta^{1/6}}{1 + \log^{1/4}(n)} \right)
\]
(50)

Here \( C \) and \( C^* \) only depend on \( c_0 \) and \( C_0 \).

We emphasize that \( \Sigma \) and \( \Sigma^* \) in lemma 4 can be singular. Theorem 3 in Chernozhukov et al. [36] holds true even if \( \Sigma \) is singular. On the other hand, if \( \Sigma^* \) is singular, from theorem 4.1.5 in Horn and Johnson [38], \( \Sigma = QQ^T \) with \( \Lambda = \text{diag}(\lambda_1, \ldots, \lambda_r, 0, 0, \ldots, 0), 0 \leq r < n \), and \( Q = \{q_{ij}\}_{i,j=1,\ldots,n} \) satisfies \( QQ^T = Q^T Q = I_n \), the n dimensional identity matrix.

Define \( \tau = Q^T \xi = (\tau_1, \ldots, \tau_r, 0, \ldots, 0)^T \) almost surely. For any continuous differentiable function such that \( E \sum_{i=1}^{n} |\partial_i f| < \infty \), lemma 2 in [37] implies

\[
E\xi f(\xi) = \sum_{j=1}^{r} q_{j} E\tau_j f(Q \tau) = \sum_{j=1}^{r} \sum_{k=1}^{n} q_{j,k} E(\tau_j \tau_k) \times E \partial_k f(Q \tau)
\]
\[= \sum_{l=1}^{n} \sigma_{ll}^* E \partial_l f(Q \tau)
\]
(51)
The validity of (51) for degenerated \( \xi \) makes sure that (50) works for degenerated \( \xi \).

**Proof of Lemma** Since \( |\epsilon_i| = \max(\epsilon_i, -\epsilon_i) \Rightarrow \max_{i=1, \ldots, n} |\epsilon_i| = \max(\max_{i=1, \ldots, n} \epsilon_i, \max_{i=1, \ldots, n} -\epsilon_i) \), and \( -\epsilon \) has the same joint distribution as \( \epsilon \),

\[
\sup_{x \in \mathbb{R}} \left( \text{Prob} \left( \max_{i=1, \ldots, n} |\epsilon_i| \leq x + \delta \right) - \text{Prob} \left( \max_{i=1, \ldots, n} |\epsilon_i| \leq x \right) \right) \\
\leq \sup_{x \in \mathbb{R}} \text{Prob} \left( x < \max_{i=1, \ldots, n} \epsilon_i \leq x + \delta \right) + \sup_{x \in \mathbb{R}} \text{Prob} \left( x < \max_{i=1, \ldots, n} -\epsilon_i \leq x + \delta \right) \\
\leq 2 \sup_{x \in \mathbb{R}} \text{Prob} \left( |\max_{i=1, \ldots, n} \epsilon_i - x| \leq \delta \right) 
\tag{52}
\]

From theorem 3 and (18), (19) in Chernozhukov et al. [46], define \( \sigma = \min_{i=1, \ldots, n} \mathbb{E} \epsilon_i^2 \) and \( \overline{\sigma} = \max_{i=1, \ldots, n} \mathbb{E} \epsilon_i^2 \),

\[
\sup_{x \in \mathbb{R}} \text{Prob} \left( |\max_{i=1, \ldots, n} \epsilon_i - x| \leq \delta \right) \leq \frac{\sqrt{2\delta}}{\sqrt{c_0}} \left( \sqrt{\log(n)} + \sqrt{\max(1, \log(c_0) - \log(\delta))} \right) \\
+ \frac{4\sqrt{2\delta}}{\sqrt{c_0}} \left( \sqrt{\log(n)} + 2 + \sqrt{\max(0, \log(c_0) - \log(\delta))} \right) \\
\leq \frac{\sqrt{2\delta}}{c_0} \left( \sqrt{\log(n)} + 1 + |\log(c_0)| + |\log(C_0)| + \sqrt{|\log(\delta)|} \right) \\
+ \frac{4\sqrt{2\delta}C_0}{c_0^2} \left( \sqrt{\log(n)} + 2 + |\log(c_0)| + |\log(C_0)| + \sqrt{|\log(\delta)|} \right) \\
\leq \left( \frac{\sqrt{2} \times (1 + |\log(c_0)| + |\log(C_0)|)}{c_0} \right) + \frac{4\sqrt{2\delta}C_0}{c_0^2} \times \left( 2 + \sqrt{|\log(c_0)| + |\log(C_0)|} \right) \times \delta \left( 1 + \sqrt{\log(n)} + \sqrt{|\log(\delta)|} \right) \tag{53}
\]

and we prove (19).

If \( \Delta = 0 \), then (50) holds true. So we assume \( \Delta > 0 \). Without loss of generality, assume \( \epsilon \) is independent of \( \xi \). Similar to Chernozhukov et al. [46], for any \( 0 \leq t \leq 1 \), define random variables \( Z_i(t) = \sqrt{t} \epsilon_i + \sqrt{1-t} \xi_i \). According to (18), (51) and theorem 2.27 in Folland [41],

\[
\mathbb{E} h_{r, \psi, x}(\epsilon_1, \ldots, \epsilon_n) - \mathbb{E} h_{r, \psi, x}(\xi_1, \ldots, \xi_n) = \frac{1}{2} \sum_{i=1}^{n} \int_{[0,1]} (1-t)^{-1/2} \mathbb{E} (\epsilon_i \times \partial_t h_{r, \psi, x}(Z_1(t), \ldots, Z_n(t))) \ dt \\
- \frac{1}{2} \sum_{i=1}^{n} \int_{[0,1]} (1-t)^{-1/2} \mathbb{E} (\xi_i \times \partial_t h_{r, \psi, x}(Z_1(t), \ldots, Z_n(t))) \ dt \\
= \frac{1}{2} \sum_{i=1}^{n} \sum_{i=1}^{n} (\sigma_{it} - \sigma_{it}^*) \int_{[0,1]} \mathbb{E} \partial_t h_{r, \psi, x}(Z_1(t), \ldots, Z_n(t)) \ dt \\
\Rightarrow |\mathbb{E} h_{r, \psi, x}(\epsilon_1, \ldots, \epsilon_n) - \mathbb{E} h_{r, \psi, x}(\xi_1, \ldots, \xi_n)| \leq \Delta \times (g_\psi \psi^2 + g_\psi \psi \tau) \tag{54}
\]

25
For any $x \in \mathbb{R}$ and given $\tau, \psi > 0$, define $t = \frac{1}{\psi} + \frac{\log(2n)}{\tau}$,

$$
Prob\left(\max_{i=1,\ldots,n} |\epsilon_i| \leq x\right) - Prob\left(\max_{i=1,\ldots,n} |\xi_i| \leq x\right)
\leq Prob\left(\max_{i=1,\ldots,n} |\epsilon_i| \leq x - t\right) + Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right) - Prob\left(\max_{i=1,\ldots,n} |\xi_i| \leq x\right)
\leq \mathbb{E} h_{r,\psi,x-t} (\epsilon_1, \ldots, \epsilon_n) - \mathbb{E} h_{r,\psi,x} (\xi_1, \ldots, \xi_n) + Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right)
$$

$$
Prob\left(\max_{i=1,\ldots,n} |\epsilon_i| \leq x\right) - Prob\left(\max_{i=1,\ldots,n} |\xi_i| \leq x\right)
\geq Prob\left(\max_{i=1,\ldots,n} |\epsilon_i| \leq x + t\right) - Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right) - Prob\left(\max_{i=1,\ldots,n} |\xi_i| \leq x\right)
\geq \mathbb{E} h_{r,\psi,x+t} (\epsilon_1, \ldots, \epsilon_n) - \mathbb{E} h_{r,\psi,x} (\xi_1, \ldots, \xi_n) - Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right)
$$

(55)

Therefore,

$$
\sup_{x \in \mathbb{R}} |Prob\left(\max_{i=1,\ldots,n} |\epsilon_i| \leq x\right) - Prob\left(\max_{i=1,\ldots,n} |\xi_i| \leq x\right)| \leq g_\ast \Delta \times (\psi^2 + \psi \tau)
+ Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right)
$$

(56)

Choose $\tau = \psi = \frac{1 + \log^{3/2}(n)}{\Delta^{1/3}}$, then

$$
0 < t = \frac{\Delta^{1/3}(1 + \log(2))}{1 + \log^{3/2}(n)} + \frac{\Delta^{1/3} \log(n)}{1 + \log^{3/2}(n)} \leq \frac{1 + \log(2)}{1 + \log^{3/2}(n)} + \frac{3}{1 + \log^{1/2}(n)} \leq 4 + \log(2)
$$

(57)

for any $n$. So $\exists$ a constant $C^\ast$ such that

$$
g_\ast \Delta (\psi^2 + \psi \tau) + Ct\left(1 + \sqrt{\log(n)} + \sqrt{\log(t)}\right) \leq C^\ast \Delta^{1/3} (1 + \log^3(n)) + C^\ast \frac{\Delta^{1/6}}{1 + \log^{1/4}(n)}
$$

(58)

and we prove (50).
B Proofs of lemmas in section 3

proof of lemma 2 For \( \zeta_{i,k}(j) = \epsilon_i(j - k)\epsilon_{i+k}(j) - E\epsilon_i\epsilon_{i+k} \). According to Cauchy - Schwarz inequality and (7),

\[
\|\zeta_{i,k} - \zeta_{i,k}(j)\|_{m/2} = \|\epsilon_i\epsilon_{i+k} - \epsilon_i(j - k)\epsilon_{i+k}(j)\|_{m/2} \\
\leq \|\epsilon_i - \epsilon_i(j - k)\|_{m} \times \|\epsilon_{i+k}\|_{m} + \|\epsilon_i(j - k)\|_{m} \times \|\epsilon_{i+k} - \epsilon_{i+k}(j)\|_{m} \\
\Rightarrow \Delta_{k,m/2} \leq C\delta_{k,m} + C\delta_{j,m}, \text{ Here } C \text{ is a constant (59)}
\]

\[
\Rightarrow (1 + j)^\alpha \sum_{l=j}^\infty \Delta_{k,l,m/2} \leq C(1 + j)^\alpha \sum_{l=j}^\infty \delta_{l-k,m} + C(1 + j)^\alpha \sum_{l=j}^\infty \delta_{l,m}
\]

and we prove (10).

On the other hand, since \( \Delta_{k,l,m/2} = 0 \) for \( l < 0 \),

\[
\|\nu_{i,t,j,k} - \nu_{i,t,j,k}(l)\|_{m/4} \\
= \|\zeta_{i,j} \zeta_{i+\ell,k} - \zeta_{i,j}(l + j - j \vee (k + \ell))\zeta_{i+\ell,k}(l + k + t - j \vee (k + \ell))\|_{m/4} \\
\leq \|\zeta_{i,j} - \zeta_{i,j}(l + j \vee (k + \ell))\|_{m/2} \times \|\zeta_{i+\ell,k}\|_{m/2} \\
+ \|\zeta_{i,j}(l + j \vee (k + \ell))\|_{m/2} \times \|\zeta_{i+\ell,k} - \zeta_{i+\ell,k}(l + k + t - j \vee (k + \ell))\|_{m/2} \ \\
\Rightarrow \Theta_{(t,t,j,k),l,m/4} \leq C\Delta_{j,l+j\vee(k+\ell),m/2} + C\Delta_{k,l+k\vee(k+\ell),m/2}, \text{ C is a constant (60)}
\]

\[
\Rightarrow \sum_{l=0}^\infty \Theta_{(t,t,j,k),l,m/4} \leq C \sum_{l=0}^\infty \Delta_{j,l,m/2} + C \sum_{l=0}^\infty \Delta_{k,l,m/2}
\]

and we prove (17).

proof of lemma 3 For any fixed \( i,j \), the random variables \( \ldots \epsilon_{l-1}, \epsilon_0, \epsilon_1, \ldots \) and the random variables \( \ldots \epsilon_{i-j-1}, \epsilon_{i-j}, \epsilon_{i-j+1}, \ldots \) have the same joint distribution. Define \( \psi_l \) as in lemma 1.

From lemma 1

\[
\Delta_{i,j,m} \leq \sum_{k=0}^\infty |\psi_k| \times \|\epsilon_{i-k} - \epsilon_{i-k}(j - k)\|_{m} \\
\Rightarrow \sum_{j=k}^{\infty} \sup_{i \in \mathbb{Z}} \Delta_{i,j,m} \leq \sum_{j=k}^{\infty} \sum_{l=0}^\infty |\psi_l| \delta_{j-l,m} \leq \sum_{l=0}^\infty |\psi_l| \times \frac{C}{(1 + (k - l) \vee 0)^\alpha} \leq \frac{C}{(1 + k/2)^\alpha} \sum_{l=0}^{[k/2]} |\psi_l| + C \sum_{l=[k/2] + 1}^\infty |\psi_l| \ \\
\text{Here } C \text{ is a constant and } [x] \text{ denotes the largest integer } a \text{ such that } a \leq x. \text{ For } |\psi_l| \leq
\]

27
$C(1 + \delta)^{-1}$ with a constant $\delta > 0$, we prove (18).

\section*{C Proofs of theorems in section 4}

\textit{proof of theorem 1.} For $\|\hat{\gamma}_i - \gamma_i\|_{m/2} \leq \|\hat{\gamma}_i - E\hat{\gamma}_i\|_{m/2} + \frac{C}{n}|\gamma_i|$, form lemma 3, lemma 2 and (14)

\begin{equation}
\|\hat{\gamma}_i - E\hat{\gamma}_i\|_{m/2} = \|\frac{1}{n} \sum_{j=1}^{n} (X_j \gamma_j - \gamma_i)\|_{m/2} \\
\leq \frac{2C\sqrt{n-1}}{n} \left(\max_{j \in \mathbb{Z}} \|X_j\|_{m} \times \max_{i \in \mathbb{Z}} \|X_j - X_{j-i}\|_{m/2} \right) (62)
\end{equation}

In particular, $\hat{\gamma}_i - \gamma_i = O_p(1/\sqrt{n})$. From assumption 5 and (5.8.4) in [38], we have $|\Gamma^{-1} - \hat{\Gamma}^{-1}|_2 = O_p(1/\sqrt{n})$. And $\hat{\Gamma}$ is non-singular with probability tending to 1. So

\begin{equation}
|\hat{\gamma}_i - \gamma_i|_2 \leq |\hat{\Gamma}^{-1} - \Gamma^{-1}|_2 \times |\hat{\gamma}_i|_2 + |\Gamma^{-1} - \hat{\Gamma}|_2 = O_p(1/\sqrt{n})
\end{equation}

and we prove (26).

\textit{proof of theorem 2.} First notice that

\begin{equation}
\max_{i=1, \ldots, p} \left| \sum_{j=0}^{p} b_{ij}(E\hat{\gamma}_j - \gamma_j) \right| = \frac{1}{n} \max_{i=1, \ldots, p} \left| \sum_{j=0}^{p} b_{ij} \times j \gamma_j \right| = O(1/n) (64)
\end{equation}

So $\exists$ a constant $C > 0$ such that

\begin{align}
\text{Prob} \left( \max_{i=1, \ldots, p} \left| \sqrt{n} \sum_{j=0}^{p} b_{ij}(\hat{\gamma}_j - \gamma_j) \right| \leq x \right) \\
\leq \text{Prob} \left( \max_{i=1, \ldots, p} \left| \sqrt{n} \sum_{j=0}^{p} b_{ij}(E\hat{\gamma}_j - \gamma_j) \right| \leq x + C/n \right) \\
\text{Prob} \left( \max_{i=1, \ldots, p} \left| \sqrt{n} \sum_{j=0}^{p} b_{ij}(\hat{\gamma}_j - \gamma_j) \right| \leq x \right) \\
\geq \text{Prob} \left( \max_{i=1, \ldots, p} \left| \sqrt{n} \sum_{j=0}^{p} b_{ij}(\hat{\gamma}_j - E\hat{\gamma}_j) \right| \leq x - C/n \right)
\end{align}
for any $n$. From assumption 5,

$$
E_{ij}^2 = \sum_{j_1=0}^p \sum_{j_2=0}^p b_{j_1} b_{j_2} \times \frac{1}{n} \sum_{j_1=j_2+1}^n \sum_{j_2=j_2+1}^n E(X_{l_1} X_{l_1-j_1} - \gamma_{j_1}) (X_{l_2} X_{l_2-j_2} - \gamma_{j_2})
$$

\[ \geq C \sum_{j=0}^p b_{j}^2 > 0 \] (66)

for sufficiently large $n$. From theorem 1

$$
\|\sqrt{n} \sum_{j=0}^p b_{j} (\gamma_j - E\gamma_j)\|_2 \leq \sum_{j=0}^p |b_{j}| \times \|\sqrt{n}(\gamma_j - E\gamma_j)\|_2 \leq C \sqrt{p+1} \times \sqrt{\sum_{j=0}^p b_{j}^2} \] (67)

so from lemma 4 and 55, for any $\tau, \psi > 0$, define $t = \frac{1}{\psi} + \frac{\log(2p)}{\tau}$.

$$
\sup_{x \in \mathbb{R}} |\text{Prob} \left( \sqrt{n} \max_{i=1, ..., p_1} \|\sum_{j=0}^p b_{ij} (\gamma_j - E\gamma_j)\|_2 \leq x \right) - \text{Prob} \left( \max_{i \in \{1, \ldots, p_1\}} |\xi_i| \leq x \right) | 
 \leq C t (1 + \sqrt{\log(p_1) + \sqrt{\log(t)}}) 
$$

$$
+ \sup_{x \in \mathbb{R}} |\text{E}_{h, \psi, \tau} \left( \sqrt{n} \sum_{j=0}^p b_{ij} (\gamma_j - E\gamma_j), \ldots, \sqrt{n} \sum_{j=0}^p b_{ij} (\gamma_j - E\gamma_j) \right) - \text{E}_{h, \psi, \tau} (\xi_1, \ldots, \xi_{p_1}) | 
$$

(68)

Here $C$ is a constant. For any integer $s \geq p + 1$, define $\epsilon_{i,j,s} = E(X_i X_{i-j} - \gamma_j)|F_{i,s}$, here $j \geq 0$. From lemma 2 lemma 3 and 11, $\exists$ a constant $C$ such that for $j = 0, 1, ..., p$ and $s \geq p + 1$

$$
\|\frac{1}{n} \sum_{i=j+1}^n (X_i X_{i-j} - \gamma_j) - \frac{1}{n} \sum_{i=j+1}^n \epsilon_{i,j,s}\|_{m/2} \leq \frac{1}{n} \sum_{i=j+1}^\infty |\epsilon_{i,j,s}^\tau| - \frac{1}{n} \sum_{i=j+1}^\infty |\epsilon_{i,j,s}^{-\tau}|\|_{m/2} 
$$

\[ \leq \sum_{t=s+1}^\infty \frac{1}{n} \sum_{i=j+1}^n X_i X_{i-j} - \frac{1}{n} \sum_{i=j+1}^n X_i(t) X_{i-j}(t-j)\|_{m/2} \leq \frac{C}{\sqrt{n} \times (1 + s)^m} \] (69)

Therefore, define $r_{j,s} = \frac{1}{n} \sum_{i=j+1}^{s+1} \epsilon_{i,j,s}$,

$$
\sup_{x \in \mathbb{R}} |\text{E}_{h, \psi, \tau} \left( \sqrt{n} \sum_{j=0}^p b_{ij} (\gamma_j - E\gamma_j), \ldots, \sqrt{n} \sum_{j=0}^p b_{ij} (\gamma_j - E\gamma_j) \right) - \text{E}_{h, \psi, \tau} \left( \sqrt{n} \sum_{j=0}^p b_{ij} r_{j,s} \right) | 
\leq g_{\psi} \sqrt{n} \times \text{E} \max_{i=1, ..., p_1} \sum_{j=0}^p |b_{ij} (\gamma_j - E\gamma_j - r_{j,s})| 
$$

\[ \leq g_{\psi} \sqrt{n} \times \sum_{i=1}^{p_1} \sum_{j=0}^n |b_{ij}| \times \|\frac{1}{n} \sum_{i=j+1}^n (X_i X_{i-j} - \gamma_j) - \frac{1}{n} \sum_{i=j+1}^n \epsilon_{i,j,s}\|_{m/2} \leq \frac{C\psi}{(1 + s)^m} \] (70)
For any $i$,
\begin{equation}
\sqrt{n} \sum_{j=0}^{p} b_{ij} r_{j,s} = \frac{1}{\sqrt{n}} \sum_{l=1}^{n} p(l-1) \sum_{j=0}^{b_{ij} t_{l,j,s}}
\end{equation}
(71)

define $L_{i,l,s} = \sum_{j=0}^{p(l-1)} b_{ij} t_{j,s}$ for $l = 1, 2, \ldots, n$. Then $L_{i,l,s}$ is $F_{l,s}$ measurable. For any given $k > s$, define the big block $S_{i,l,s} = \frac{1}{\sqrt{n}} \sum_{l=1}^{(l-1) \times (k+s+1)} L_{i,l,s}$ and the small block $s_{i,l,s} = \frac{1}{\sqrt{n}} \sum_{l=(l-1) \times (k+s)+1}^{(l-1) \times (k+s)+k+1} L_{i,l,s}$, where $l = 1, 2, \ldots, R$, $R = \lceil \frac{n}{k+s} \rceil$. $[x]$ denotes the smallest integer that is larger than or equal to $x$. Then the random vectors $(S_{1,l,s}, \ldots, S_{p,l,s})^T$, $l = 1, 2, \ldots, R$ are mutually independent; and the random vectors $(s_{1,l,s}, \ldots, s_{p,l,s})^T$ are mutually independent. We also have
\begin{equation}
\sqrt{n} \sum_{j=0}^{p} b_{ij} r_{j,s} = \sum_{l=1}^{R} S_{i,l,s} + \sum_{l=1}^{R} s_{i,l,s}
\end{equation}
(72)

From theorem 2 in [35],
\begin{equation}
|Es_{i,l,s} \left( \left| \sqrt{n} \sum_{j=0}^{p} b_{ij} r_{j,s} \right| \right) | \leq g \psi \psi_{i,l,s} \sum_{l=1}^{R} \left| s_{i,l,s} \right| \leq g \psi \psi \sum_{l=1}^{p} \left| s_{i,l,s} \right|_{m/2} = O \left( \psi_{i,l,s} \psi_{l,s} \sum_{l=1}^{R} \left| s_{i,l,s} \right|_{m/2} \right)
\end{equation}
(73)

Define $\Delta_{i,l,s,j} = E L_{i,l,s} | F_{l,j} - E L_{i,l,s} | F_{l,j-1}$, we have
\begin{equation}
L_{i,l,s} = E L_{i,l,s} | F_{l,0} + \sum_{j=1}^{s} (E L_{i,l,s} | F_{l,j} - E L_{i,l,s} | F_{l,j-1}) = E L_{i,l,s} | F_{l,0} + \sum_{j=1}^{s} \Delta_{i,l,s,j}
\end{equation}
(74)

and $\Delta_{i,l,s,j} = \sum_{i=0}^{p(l-1)} b_{i,j} (E X_{i} X_{i-j} | F_{l,j} - E X_{i} X_{i-j} | F_{l,j-1})$ for $j \leq s$

From (10)
\begin{equation}
\|L_{i,l,s}\|_{m/2} \leq \sum_{j=0}^{p(l-1)} \|b_{i,j}\| \times \|X_{i} X_{i-j}\|_{m/2} \Rightarrow \sup_{l \in \mathbb{Z}, s \geq 0} \|L_{i,l,s}\|_{m/2} < \infty
\end{equation}
(75)

and
\begin{equation}
\|\Delta_{i,l,s,j}\|_{m/2} \leq \sum_{i=0}^{p(l-1)} \|b_{i,j}\| \times \|X_{i} X_{i-j} - X_{i}(j) X_{i-j}(j-v)\|_{m/2} \leq C \sum_{i=0}^{p} \sup_{l \in \mathbb{Z}} \|X_{i} - X_{i}(j-v)\|_{m/2}
\end{equation}
(76)
for a constant $C$. Notice that

$$
\|s_{i,j,s}\|_{m/2} \leq \frac{1}{\sqrt{n}} \sum_{z = ((l+1) \times (k+1)) \cap n}^{((l+1) \times (k+1)) \cap n} E \|L_{i,z,s}\|_{m/2} + \sum_{j=1}^{a} \frac{1}{\sqrt{n}} \sum_{z = ((l-1) \times (k+1)) \cap n} \Delta_{i,z,s,j}\|_{m/2}
$$

(77)

from theorem 2 in Whittle [35]

$$
\frac{1}{\sqrt{n}} \sum_{z = ((l-1) \times (k+1)) \cap n}^{((l+1) \times (k+1)) \cap n} E \|L_{i,z,s}\|_{m/2} \leq \frac{C \sqrt{s}}{\sqrt{n}} \sup_{z \in \mathbb{Z}} \|L_{i,z,s}\|_{m/2}
$$

(78)

from (12).

$$
\frac{1}{\sqrt{n}} \sum_{z = ((l-1) \times (k+1)) \cap n} \Delta_{i,z,s,j}\|_{m/2} \leq \frac{C \sqrt{s}}{\sqrt{n}} \sup_{z \in \mathbb{Z}} \|\Delta_{i,z,s,j}\|_{m/2}
$$

(79)

Therefore, from lemma 2 and 3.

$$
|E_h_{i,j,s} \left( \sqrt{n} \sum_{j=0}^{p} b_{ij} r_{j,s}, ..., \sqrt{n} \sum_{j=0}^{p} b_{ij} r_{j,s} \right) - E_h_{i,j,s} \left( \sum_{i=1}^{R} (S_{1,i}, ..., \sum_{l=1}^{R} S_{l,i,s} \right) | = O \left( \psi \sqrt{\frac{s}{k}} \right)
$$

(80)

Define $S_{j,s} = (S_{1,j,s}, ..., S_{p_1,j,s})^T, j = 1, ..., R$ as joint normal random vectors with $E S_{j,s} = 0$ and $E S_{1,j,s} S_{2,j,s} = ES_{1,j,s} S_{2,j,s}^T$. $S_{j,s}$ is independent with $S_{j,s}^T$ for $j \neq k$; and $S_{j,s}$ is independent with $(S_{1,k,s}, ..., S_{p_1,k,s})^T$ for any $k$. Here $i, i_1, i_2 = 1, ..., p_1$. Define $H_{i,j,s} = \sum_{k=1}^{j-1} S_{i,k,s} + \sum_{k=j+1}^{R} S_{i,k,s}, \text{ then } H_{i,j,s} + S_{i,j,s} = H_{i,j+1,s} + S^{*}_{i,j+1,s}$. Define $H_{i,s} = (H_{1,i,s}, ..., H_{p_1,i,s})^T$ and $S_{i,s} = (S_{1,i,s}, ..., S_{p_1,i,s})^T$. From Taylor’s theorem and [35]

$$
|E \left( h_{r,v,s} (H_{i,s} + S_{j,s}) - h_{r,v,s} (H_{i,s} + S_{j,s}) \right) | H_{i,s} |
$$

$$
\leq |E \left( \sum_{i=1}^{p_1} \partial_i h_{r,v,s} (H_{i,s}) (S_{i,j,s} - S_{i,j,s}^*) \right) | H_{i,s} |
$$

$$
+ \frac{1}{2} |E \left( \sum_{i=1}^{p_1} \partial_i h_{r,v,s} (H_{i,s}) (S_{1,i,s}) (S_{i,j,s} - S_{i,j,s}^*) (S_{1,i,s} - S_{i,j,s}^*) \right) | H_{i,s} |
$$

$$
+ g_{\tau} (\psi^3 + 6\psi^2 \tau + 3\psi^2 \tau) \times \left( E \max_{i=1, ..., p_1} |S_{i,j,s}|^3 + E \max_{i=1, ..., p_1} |S_{i,j,s}^*|^3 \right)
$$

(81)

$$
\leq g_{\tau} (\psi^3 + \psi^2 \tau + \psi^2 \tau) \sum_{i=1}^{p_1} (\|S_{i,j,s}\|_3^3 + \|S_{i,j,s}^*\|_3^3)
$$
So

\[
\sup_{x \in \mathbb{R}} |E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) - E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s} \right) | \leq \sup_{x \in \mathbb{R}} \left| E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) - E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) \right| = O \left( \psi^3 + \psi^2 \tau + \psi \tau^2 \right) \sum_{j=1}^{R} \sum_{i=1}^{p_1} (\|S_{i,j,s}\|_3^3 + \|S_{i,j,s}^*\|_3^3) \tag{82}
\]

For \(S_{i,j,s}^*\) has normal distribution, \(\exists\) a constant \(C\) such that \(\|S_{i,j,s}^*\|_3 \leq C\|S_{i,j,s}\|_3\) \(\leq C\|S_{i,j,s}\|_3^2\). Similar to (77),

\[
\|S_{i,j,s}\|_3 \leq \|S_{i,j,s}\|_{m/2} \leq C \sqrt{k/n} \tag{83}
\]

for any \(j\). Here \(C\) is a constant. From section A,

\[
E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) - E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s} \right) \]

\[
\leq \text{Prob} \left( \max_{i=1, \ldots, p_1} \sum_{j=1}^{R} S_{i,j,s}^* \leq x + \frac{1}{\psi} \right) - \text{Prob} \left( \max_{i=1, \ldots, p_1} |\xi_i| \leq x - \frac{\log(2p_1)}{\tau} \right) \tag{84}
\]

\[
E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) - E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s} \right) \]

\[
\geq \text{Prob} \left( \max_{i=1, \ldots, p_1} \sum_{j=1}^{R} S_{i,j,s}^* \leq x - \frac{\log(2p_1)}{\tau} \right) - \text{Prob} \left( \max_{i=1, \ldots, p_1} |\xi_i| \leq x + \frac{1}{\psi} \right)
\]

\[
\Rightarrow \sup_{x \in \mathbb{R}} \left| E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s}^* \right) - E_{h, \tau, \psi, x} \left( \sum_{j=1}^{R} S_{j,s} \right) \right| \leq \sup_{x \in \mathbb{R}} \left| \text{Prob} \left( \max_{i=1, \ldots, p_1} \sum_{j=1}^{R} S_{i,j,s}^* \leq x \right) - \text{Prob} \left( \max_{i=1, \ldots, p_1} |\xi_i| \leq x \right) \right| + \sum_{i=1}^{p_1} \sup_{x \in \mathbb{R}} \text{Prob} \left( x - \frac{\log(2p_1)}{\tau} \leq |\xi_i| \leq x + \frac{1}{\psi} \right)
\]

Since \(\sqrt{n} \sum_{j=0}^{p} b_{ij} (\gamma_j - E\gamma_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sum_{j=0}^{p} b_{ij} (X_i \gamma_{i-1} - \gamma_j)\), define \(T_{i,t} = \sum_{j=0}^{p} b_{ij} (X_i X_{i-1} - \gamma_j)\)....
\(\gamma_l\) for \(l = 1, 2, ..., n\). Since \(\mathbf{ES}_{i_1,j_1,s}^* S_{i_2,j_2,s}^* = \mathbf{ES}_{i_1,j_1,s} S_{i_2,j_2,s} = 0\) if \(j_1 \neq j_2\), we have
\[
\left| \sum_{j_1=1}^{R} \sum_{j_2=1}^{R} \mathbf{ES}_{i_1,j_1,s}^* S_{i_2,j_2,s}^* - \mathbf{ES}_{i_1,j_1,s} S_{i_2,j_2,s} \right| = \left| \sum_{j_1=1}^{R} \sum_{j_2=1}^{R} \mathbf{ES}_{i_1,j_1,s} S_{i_2,j_2,s} - \frac{1}{n} \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} \mathbf{ET}_{i_1,l_1} T_{i_2,l_2} \right|
\leq \left| \sum_{j_1=1}^{R} \sum_{j_2=1}^{R} \mathbf{ES}_{i_1,j_1,s} S_{i_2,j_2,s} - \mathbf{E} \left( \sum_{l_1=1}^{R} S_{i_1,l_1,s} + \sum_{l_1=1}^{R} s_{i_1,l_1,s} \right) \right| \times \left( \sum_{l_1=1}^{R} S_{i_2,l_1,s} + \sum_{l_1=1}^{R} s_{i_2,l_1,s} \right) + \frac{1}{n} \sum_{l_1=1}^{n} \sum_{l_2=1}^{n} \left( \mathbf{E} L_{i_1,l_1,s} L_{i_2,l_2,s} - \mathbf{ET}_{i_1,l_1} T_{i_2,l_2} \right)
\leq \left\| \sum_{j=1}^{R} S_{i_1,j,s} \right\|_2 \times \left\| \sum_{j=1}^{R} S_{i_2,j,s} \right\|_2 + \left\| \sum_{j=1}^{R} S_{i_1,j,s} + \sum_{j=1}^{R} s_{i_1,j,s} \right\|_2 \times \left\| \sum_{j=1}^{R} S_{i_2,j,s} + \sum_{j=1}^{R} s_{i_2,j,s} \right\|_2 + \frac{1}{n} \left\| \sum_{l=1}^{n} L_{i_1,l,s} \right\|_2 \times \left\| \sum_{l=1}^{n} T_{i_2,l} \right\|_2 + \frac{1}{n} \left\| \sum_{l=1}^{n} T_{i_1,l} \right\|_2 \times \left\| \sum_{l=1}^{n} (L_{i_1,l,s} - T_{i_1,l}) \right\|_2
\]

(85)

For \(L_{i_1,l,s} = \mathbf{ET}_{i_1,l} |\mathbf{F}_{l,s}\), from (14) and lemma 2,
\[
\left\| \sum_{l=1}^{n} (L_{i_1,l,s} - T_{i_1,l}) \right\|_2 \leq \sum_{l=1}^{n} \left\| L_{i_1,l,0} \right\|_m/2 + \sum_{l=1}^{n} \left\| (L_{i_1,l,k} - L_{i_1,l,k-1}) \right\|_m/2 = O(\sqrt{n})
\]

(86)

According to (74) to (79) and (83)
\[
\left\| \sum_{j=1}^{R} S_{i_1,j,s} \right\|_2^2 = \sum_{j=1}^{R} \mathbf{ES}_{i_1,j,s}^2 = O(1); \quad \left\| \sum_{j=1}^{R} s_{i_1,j,s} \right\|_2^2 = \sum_{j=1}^{R} \mathbf{Es}_{i_1,j,s}^2 = O \left( \frac{n}{R} \right)
\]

(87)

From lemma 4,
\[
\left| \sum_{j_1=1}^{R} \sum_{j_2=1}^{R} \mathbf{ES}_{i_1,j_1,s}^* S_{i_2,j_2,s}^* - \mathbf{ES}_{i_1,j_1,s} S_{i_2,j_2,s} \right| = \sup_{x \in \mathbb{R}} |\text{Prob} \left( \max_{j=1,\ldots,n_1} \sum_{j=1}^{R} S_{i,j,s} \leq x \right) - \text{Prob} \left( \max_{j=1,\ldots,n_1} |\xi_j| \leq x \right) | = O \left( \frac{\sqrt{n}}{k} + \frac{1}{(1 + s)^{\alpha_x}} \right)
\]

(88)

From assumption 5, \(\text{Prob} \left( x - \frac{\log(2n)}{n} \leq |\xi_j| \leq x + \frac{1}{n} \right) = O \left( \frac{1}{k} + \frac{1}{n} \right)\) and \(\sup_{x \in \mathbb{R}} |H(x) - H(x - \frac{\log(2n)}{n})| \leq \sum_{i=1}^{n_1} \text{Prob} \left( x - \frac{\log(2n)}{n} < |\xi| \leq x \right) = O(1/n)\).

Select \(k = \lfloor \sqrt{n} \rfloor, \ s = [n^{1/8}]\) and \(\tau = \psi = \log^4(n) \Rightarrow t = O(1/ \log^4(n))\). From (65), (68), (70), (80), (82) and (83), we prove (20).
D Proofs of theorems in section 5

This section justifies the validity of bootstrap algorithm 1. We first introduce a lemma.

Lemma 5

Suppose assumption 1 to 6. Define \( \tilde{\zeta}_{i,k} = X_i X_{i-k} - \tilde{\gamma}_k \) and \( \zeta_{i,k} = X_i X_{i-k} - \gamma_k \) for \( k = 0,1,\ldots,p \) and \( i = k+1,\ldots,n \). \( \tilde{\gamma}_k \) is defined in \([34]\). Then for any given \( k_1, k_2 = 0,1,2,\ldots,p \),

\[
\frac{1}{n} \sum_{j_1=k_1+1}^{n} \sum_{j_2=k_2+1}^{n} \tilde{\zeta}_{j_1,k_1} \tilde{\zeta}_{j_2,k_2} K\left( \frac{j_1-j_2}{k_n} \right) - \frac{1}{n} \sum_{j_1=k_1+1}^{n} \sum_{j_2=k_2+1}^{n} E \zeta_{j_1,k_1} \zeta_{j_2,k_2} \right| = O_p\left( \frac{k_n}{\sqrt{n}} + v_n \right)
\]

(89)

Here

\[
v_n = \begin{cases} 
  k_1^{1-\alpha} & \text{if } 1 \leq \alpha < 2 \\
  \log(k_n)/k_n & \text{if } \alpha = 2 \\
  1/k_n & \text{if } \alpha > 2
\end{cases}
\]

proof of lemma Define \( \tilde{\zeta}_{i,k} = \zeta_{i,k} = 0 \) for \( i = 1,2,\ldots,k \). Then for any \( i,t > 0 \) such that \( i + t \leq n \)

\[
|E \zeta_{i,k_1} \zeta_{i+t,k_2}| = |E \zeta_{i,k_1} \times (\zeta_{i+t,k_2} - E \zeta_{i+t,k_2} | F_{i+t-1})| 
\leq \|\zeta_{i,k_1}\|_2 \times \sum_{j=t}^{\infty} \|E \zeta_{i+t,k_2} | F_{i+t-j} - E \zeta_{i+t,k_2} | F_{i+t,j-1}\|_2 \leq \frac{C}{(1+t)^{\alpha}}
\]

(91)

for a constant \( C \). From section 0.9.7 in \([38]\)

\[
\frac{1}{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E \zeta_{j_1,k_1} \zeta_{j_2,k_2} K\left( \frac{j_1-j_2}{k_n} \right) - \frac{1}{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} E \zeta_{j_1,k_1} \zeta_{j_2,k_2} \right|
\leq \frac{C}{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \frac{1}{(1+|j_1-j_2|)^{\alpha}} \times \left( 1 - K\left( \frac{j_1-j_2}{k_n} \right) \right)
\]

(92)

\[
\leq 2C \sum_{s=0}^{\infty} \frac{1}{(1+s)^{\alpha}} \times \left( 1 - K\left( \frac{s}{k_n} \right) \right)
\]

\( K \) is continuous differentiable, so

\[
\sum_{s=0}^{\infty} \frac{1}{(1+s)^{\alpha}} \times \left( 1 - K\left( \frac{s}{k_n} \right) \right) \leq \max_{x \in [0,1]} |K'(x)| \sum_{s=0}^{k_n} \frac{s}{(1+s)^{\alpha}} + \sum_{s=k_n+1}^{\infty} \frac{1}{(1+s)^{\alpha}}
\]

\[
= O\left( \frac{1}{k_n} + \frac{1}{k_n} \int_{[1,k_n]} x^{1-\alpha} dx + \int_{[k_n,\infty]} x^{-\alpha} dx \right)
\]

(93)

34
Here $C, C'$ are two constants. So

$$\| \frac{1}{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} (\zeta_{j_1,1} \zeta_{j_2,2} - E\zeta_{j_1,1} \zeta_{j_2,2}) K \left( \frac{j_1 - j_2}{k_n} \right) \|_{m/4} = O \left( \frac{1}{\sqrt{n}} \sum_{l=0}^{\infty} K \left( \frac{l}{k_n} \right) \right)$$

(96)}

Since $\sum_{l=0}^{\infty} K \left( \frac{l}{k_n} \right) = O(1 + k_n \int_{0, \infty} K(x) dx) = O(k_n)$, we have

$$\| \frac{1}{n} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} (\zeta_{j_1,1} \zeta_{j_2,2} - E\zeta_{j_1,1} \zeta_{j_2,2}) K \left( \frac{j_1 - j_2}{k_n} \right) - \frac{1}{n} \sum_{j_1=b_1+1}^{n} \sum_{j_2=b_2+1}^{n} E\zeta_{j_1,1} \zeta_{j_2,2} \| = O_p \left( \frac{k_n}{\sqrt{n}} + v_n \right)$$

(97)
For
\[
\frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n \hat{\zeta}_{j_1,k_1} \hat{\zeta}_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) - \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n \zeta_{j_1,k_1} \zeta_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) \]
\leq \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n \zeta_{j_1,k_1} (\hat{\gamma}_{k_2} - \gamma_{k_2}) K \left( \frac{j_1 - j_2}{k_n} \right) + \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n (\hat{\gamma}_{k_1} - \gamma_{k_1}) \zeta_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) + \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n (\hat{\gamma}_{k_1} - \gamma_{k_1}) \times (\hat{\gamma}_{k_2} - \gamma_{k_2}) K \left( \frac{j_1 - j_2}{k_n} \right) \]
\leq \frac{2(\hat{\gamma}_{k_2} - \gamma_{k_2})}{\sqrt{n}} \sqrt{\sum_{j=1}^n \zeta_{j,k_1}^2 \sum_{k=0}^\infty K \left( \frac{s}{k_n} \right) \sum_{j=1}^n \zeta_{j,k_2}^2 \sum_{k=0}^\infty K \left( \frac{s}{k_n} \right) + 2|\hat{\gamma}_{k_1} - \gamma_{k_1}| \times (\hat{\gamma}_{k_2} - \gamma_{k_2}) \sum_{k=0}^\infty K \left( \frac{s}{k_n} \right) \sum_{j=1}^n \zeta_{j,k_1} \zeta_{j,k_2} K \left( \frac{j_1 - j_2}{k_n} \right)}
\tag{98}
\]
Since \( E \sum_{j=1}^n \zeta_{j,k_1}^2 \leq Cn \) for a constant \( C \), from theorem \[\ref{lemma5}\] we have
\[
\left| \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n \hat{\zeta}_{j_1,k_1} \hat{\zeta}_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) - \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n \zeta_{j_1,k_1} \zeta_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) \right| = O_p \left( \frac{k_n}{\sqrt{n}} \right)
\tag{99}
\]
From \[\text{\ref{eq:97}}\] and \[\text{\ref{eq:99}}\], we prove \[\text{\ref{eq:98}}\]. \( \square \)

In particular, from lemma \[\ref{lemma5}\] for given numbers \( b_{ij}, i = i_1, i_2, j = 0, 2, ..., p, \)
\[
\frac{1}{n} \sum_{k_1=0}^p \sum_{k_2=0}^p \sum_{j_1=1}^n \sum_{j_2=1}^n \hat{\zeta}_{j_1,k_1} \hat{\zeta}_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right) - \frac{1}{n} \sum_{k_1=0}^p \sum_{k_2=0}^p \sum_{j_1=1}^n \sum_{j_2=1}^n \zeta_{j_1,k_1} \zeta_{j_2,k_2} K \left( \frac{j_1 - j_2}{k_n} \right)
\tag{100}
\]
has order \( O_p \left( \frac{k_n}{\sqrt{n}} + v_n \right) \).

**proof of theorem \[\ref{thm:main}\]** Form lemma \[\ref{lemma5}\]
\[
n E^* (\hat{\gamma}_i^* - \gamma_i)^2 = \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n (X_{j_1,j_2} - \hat{\gamma}_{i}) \times (X_{j_2,j_1} - \gamma_{i}) K \left( \frac{j_1 - j_2}{k_n} \right) = \frac{1}{n} \sum_{j_1=1}^n \sum_{j_2=1}^n (X_{j_1,j_2} - \gamma_{i}) (X_{j_2,j_1} - \gamma_{i}) + O_p \left( \frac{k_n}{\sqrt{n}} + v_n \right)
\tag{101}
\]
v_n is defined in \[\text{\ref{eq:98}}\]. So for any given \( 0 < \delta < 1 \), there exists a constant \( C \) such that \( Prob^* (|\hat{\gamma}_i^* - \gamma_i| > C/\sqrt{n}) < \delta \) with probability tending to 1. From assumption 5 and theorem \[\ref{thm:main}\] the event \( \exists \ a \ constant \ c > 0 \ such \ that \ the \ smallest \ eigenvalue \ of \ \hat{\Gamma} \ is \ greater \ than \ c \) has probability tending to 1.
If \( \hat{\Gamma} \)'s smallest eigenvalue is greater than \( c \) and \(|\hat{\Gamma}^{-1}(\hat{\gamma}_i^*-\hat{\Gamma})|_2 < 1/2, \) from corollary 5.6.16
in \[ \mathbb{R}^n \], \( \hat{\Gamma}^* \) is non-singular and \( \hat{\Gamma}^{* -1} = (\sum_{k=0}^\infty (-1)^k (\hat{\Gamma}^* - \hat{\Gamma}))^k \hat{\Gamma}^{* -1} \). So

\[
\sqrt{n} (B\hat{\Gamma}^* - B\hat{\gamma}) - \left(-B\hat{\Gamma}^{* -1}(\hat{\Gamma}^* - \hat{\Gamma})\hat{\Gamma}^{* -1}\hat{\gamma} + B\hat{\Gamma}^{* -1}(\hat{\gamma}^* - \hat{\gamma})\right) |_2 \\
\leq \sqrt{n} |B||x| |\hat{\gamma}| |\hat{\Gamma}^{* -1}| | \sum_{k=2}^\infty |\hat{\Gamma}^{* -1}(\hat{\Gamma}^* - \hat{\Gamma})|_2^k + \sqrt{n} |B||x| |\hat{\Gamma}^{* -1}|^2 | \hat{\Gamma}^* - \hat{\Gamma} |_2 | |\hat{\gamma} - \hat{\gamma}|_2 \\
(102)
\]

and \( \forall \delta > 0, \exists \) a constant \( C > 0 \) such that

\[
Prob^* \left( \sqrt{n} (B\hat{\Gamma}^* - B\hat{\gamma}) - \left(-B\hat{\Gamma}^{* -1}(\hat{\Gamma}^* - \hat{\Gamma})\hat{\Gamma}^{* -1}\hat{\gamma} + B\hat{\Gamma}^{* -1}(\hat{\gamma}^* - \hat{\gamma})\right) |_2 > C/\sqrt{n} \right) < \delta \\
(103)
\]

with probability tending to 1.

Define the \( p \times p \) matrix \( T_i = \{t_{ij,k}\}, k=1,...,p \) such that \( t_{ij,k} = 1 \) if \( j-k = i \), and 0 otherwise. And define \( \delta_i = (0,0,...,0,1,0,...,0)^T \). Then

\[
-\sqrt{n} B\hat{\Gamma}^{* -1}(\hat{\Gamma}^* - \hat{\Gamma})\hat{\Gamma}^{* -1}\hat{\gamma} + \sqrt{n} B\hat{\Gamma}^{* -1}(\hat{\gamma}^* - \hat{\gamma}) \\
= \sum_{i=1}^p \sqrt{n}(\hat{\gamma}_p^* - \hat{\gamma}_p) B\hat{\Gamma}^{* -1}\delta_p - \sum_{i=1}^{p-1} \sqrt{n}(\hat{\gamma}_i^* - \hat{\gamma}_i) B\hat{\Gamma}^{* -1}T_i\hat{\Gamma}^{* -1}\hat{\gamma} \\
= \sum_{j=1}^p \sqrt{n}(\hat{\gamma}^*_j - \hat{\gamma}_j) \times \left( b_j^T B\hat{\Gamma}^{* -1}\delta_j - b_j^T B\hat{\Gamma}^{* -1}(T_j + T_{-j})\hat{\Gamma}^{* -1}\hat{\gamma} \right) + A_i \\
= \sum_{j=0}^p \tilde{c}_j \sqrt{n}(\hat{\gamma}^*_j - \hat{\gamma}_j) + A_i \\
(104)
\]

Here \( \tilde{c}_0 = -b_0^T B\hat{\Gamma}^{* -1}T_0\hat{\Gamma}^{* -1}\hat{\gamma} \), \( \tilde{c}_p = b_p^T B\hat{\Gamma}^{* -1}\delta_p \), and \( \tilde{c}_j = b_j^T B\hat{\Gamma}^{* -1}\delta_j - b_j^T B\hat{\Gamma}^{* -1}(T_j + T_{-j})\hat{\Gamma}^{* -1}\hat{\gamma} \) for \( j = 1, 2, ..., p - 1 \). From theorem 1 \( |c_{ij} - c_{ij}| = O_p(1/\sqrt{n}) \). \( c_{ij} \) is defined in corollary 1. From (103) \( A \) satisfies \( Prob^* (|A_i| > C/\sqrt{n}) < \delta \) with probability tending to 1. From lemma 2 \( \sqrt{n} \) for any \( i_1, i_2 \),

\[
n E^* \sum_{j_1=0}^p \tilde{c}_{i_1,j_1}(\hat{\gamma}^*_{j_1} - \hat{\gamma}_{j_1}) \times \sum_{j_2=0}^p \tilde{c}_{i_2,j_2}(\hat{\gamma}^*_{j_2} - \hat{\gamma}_{j_2}) = n \sum_{j_1=0}^p \sum_{j_2=0}^p \tilde{c}_{i_1,j_1} \tilde{c}_{i_2,j_2} E^*(\hat{\gamma}^*_{j_1} - \hat{\gamma}_{j_1}) \times (\hat{\gamma}^*_{j_2} - \hat{\gamma}_{j_2}) \\
= \frac{1}{n} \sum_{j_1=0}^p \sum_{j_2=0}^p \tilde{c}_{i_1,j_1} \tilde{c}_{i_2,j_2} \sum_{l_1=0}^n \sum_{l_2=0}^n \sum_{j_1=0}^p \sum_{j_2=0}^p \sum_{l_1=0}^n \sum_{l_2=0}^n \sum_{l_2=0}^n (X_{l_1}X_{l_1-j_1} - \hat{\gamma}_{j_1})(X_{l_1}X_{l_2-j_2} - \hat{\gamma}_{j_2}) \times K \left( \frac{l_1 - l_2}{k_n} \right) \\
= \frac{1}{n} \sum_{j_1=0}^p \sum_{j_2=0}^p \sum_{l_1=0}^n \sum_{l_2=0}^n E(X_{l_1}X_{l_1-j_1} - \hat{\gamma}_{j_1})(X_{l_1}X_{l_2-j_2} - \hat{\gamma}_{j_2}) + O_p(k_n/\sqrt{n} + v_n) \\
(105)
\]

37
From lemma 4 we prove (42).