Consistent bootstrap one-step prediction region for state-vector in state space model

Yunyi Zhang, Tingting Wang, Dimitris N. Politis

July 19, 2019

Abstract

In this paper, we propose a bootstrap algorithm to construct non-parametric prediction region for a simplified state-space model and provide theoretical proof for its consistency. Besides, we introduce prediction problem under the situation that innovation depends on previous observations and extend definition in [1] to a broader class of models. Numerical results show that performance of proposed bootstrap algorithm depends on smoothness of data and deviation of last observation and state variable. Results in this paper can be adjusted to other classical models, like stochastic volatility model, exogenous time series model and etc.

1 Introduction and frequently used assumptions

1.1 Introduction

Statistical inference generally consists of two aspects, to explain a model (estimation) and to predict future state of a model (prediction) [2]. Estimation is a traditional task for statistician but prediction, and especially non-parametric prediction problem also plays a more and more important role in regression and time series analysis. In regression setting, if the innovations obey normal distribution, we refer chapter 5.3 of [3] for an introduction of prediction problem. On the other hand, if the distribution of innovations is not known, bootstrap algorithm can be applied to construct prediction region, like [4] does.

For dependent data, prediction problem becomes more tricky and requires more clarification. Unlike independent case, suppose the data we gather are $x_1, ..., x_n$, constructing $1 - \alpha$ prediction region means to find a set $A = A(x_1, ..., x_n)$ such that conditional probability $Prob(x_{n+1} \in A|x_1, ..., x_n) \approx 1 - \alpha$ for sufficiently large sample size $n$. If innovation being independent on previous observations, like linear and non-linear autoregressive model (see (6.3) and (7.1) in [2] as examples), then conditional probability becomes unconditional one and prediction in this setting is well-defined, see [1] for a detail discussion. In this setting, residue-based bootstrap can be applied to construct prediction regions, like [5] for autoregressive model, [6] for joint prediction region in autoregressive model and [1] for prediction in non-linear time series.
A more complicated situation is when innovations depend on previous observations. An example for this situation is state space model

\[ X_{n+1} = AX_n + \epsilon_{n+1}, \quad Y_n = X_n + \eta_n, \quad \epsilon_n \sim i.i.d.(0, \Sigma), \quad \eta_n \sim i.i.d.(0, \Xi) \tag{1} \]

with innovations \( \epsilon_n, n = ..., -1, 0, 1, ... \) being independent of errors \( \eta_n, n = ..., -1, 0, 1, ... \) and the observed data are \( Y_1, Y_2, ..., Y_n \). Because we cannot observe data \( X_i, i = 1, 2, ..., n \) and coefficient matrix \( A \), the \( (n+1) \)th residue \( \xi_n \) and predictive root \( \hat{\xi}_n \) are actually

\[ \xi_n = Y_{n+1} - AY_n = \epsilon_{n+1} + \eta_{n+1} - A\eta_n, \quad \hat{\xi}_n = Y_{n+1} - \hat{A}Y_n = \xi_n + (A - \hat{A})Y_n \tag{2} \]

which depends on previous observation \( Y_n = X_n + \eta_n \). In this situation, as we will see in section 4, residue-based bootstrap does not work for model (1). Another example is prediction problem for Markov process which is introduced in [7].

We are interested in non-parametric prediction for state vector in the simplified state space model (1). State space model and Kalman filter are fundamental tools in system engineering [8] and time series analysis [9], yet majority researches concentrate on parametric and non-parametric estimation ([11], [12] and [13]), Kalman-filter based prediction and interpolation ([10] and [14]), modelling time series through state space model [15] and etc. If the innovations \( \epsilon_i \) and errors \( \eta_i, i = ..., -1, 0, 1, ... \) in model (1) are normally distributed, Kalman-filter can be applied to generate predictor and solve prediction problem introduced in chapter 12.2, [10]. However, in non-parametric setting, predictor generated by Kalman-filter may not be optimal. Worse still, Kalman filter cannot be applied to construct prediction regions and there are relatively few researches on non-parametric prediction problem for state space model.

In this paper, we aims at providing a consistent bootstrap algorithm for constructing prediction region of state vector under model (1). Conditional distribution of state vector is a function of previous observations rather than a constant term, which influences definition of consistency. Thus, we will first introduce prediction with dependent data and provide definition and a sufficient condition for consistency under model (1). We then extend multivariate deconvolution results in [16] to weakly dependent data. Finally, we provide a bootstrap algorithm for prediction and prove its consistency.

The structure of this paper is as follow: assumptions and notations are given in section 1.2. In section 2 we will discuss prediction for dependent data and generalize the definition and sufficient conditions given by Pan and Politis [1] to adjust the situation when innovation depends on last observations. A Bootstrap algorithm for prediction problem of state space model as well as its theoretical justification will be given in section 3. In section 4 we will perform numerical experiments discuss finite sample behavior of bootstrap algorithm. Conclusion will be given in section 5 and proof details will be given in appendix (section 6).
1.2 Notations and assumptions

In this paper, we will use three types of norms and we will define all of them here. Suppose the observed data \( Y_1, \ldots, Y_n \in \mathbb{R}^d \), we define norm \( \| \cdot \| \) being a vector norm in \( \mathbb{R}^d \) and \( \| \cdot \|_2 \) being Euclidean norm \( \| x \|_2 = \sqrt{x_1^2 + \cdots + x_d^2} \) for \( x = (x_1, \ldots, x_d) \in \mathbb{R}^d \). For a \( d \times d \) matrix \( A \), we define its two norm as

\[
\| A \|_2 = \sup_{\| x \|_2 = 1} \| Ax \|_2
\]

(3)

For a function \( f : \mathbb{R}^d \to \mathbb{C} \), we define its \( L_p \) norm (\( p \geq 1 \)) as

\[
\| f \|_{L_p} = \left( \int_{\mathbb{R}^d} |f|^p \, dx \right)^{1/p}, \quad p < \infty, \quad \| f \|_{L_\infty} = \text{ess sup} |f|
\]

(4)

Here the integration is with respect to Lebesgue measure \( m(.) \) in \( \mathbb{R}^d \). The frequently used \( L_p \) norm is when \( p = 1, 2, \infty \). Detail explanation about \( L_p \) norm and essential supremum can be seen in section 6.1 of [17].

For a random variable \( X \), we can also define its \( p \) norm associated with probability space \( \Omega \), we use notation \( \| \cdot \|_{R_p} \) to denote \( p \) norm in probability space \( \Omega \)

\[
\| X \|_{R_p} = \left( \mathbb{E} |X|^p \right)^{1/p}
\]

(5)

Here \( \mathbb{E} \) denotes expectation.

Consider stationary state space model

\[
X_{n+1} = AX_n + \epsilon_{n+1}, \quad Y_n = X_n + \eta_n, \quad X_n \in \mathbb{R}^d, \quad \| A \| < 1
\]

(6)

with \( A \) being full rank matrix satisfying \( \| A \|_2 < 1 \), \( \epsilon_n, \eta_n, n \in \mathbb{Z} \) are i.i.d random variable and their distribution being absolutely continuous with Lebesgue measure. Besides, \( \eta_k, \epsilon_j \) being independent for any \( k, j \in \{\ldots, -1, 0, 1, \ldots\} \). We suppose that \( \mathbb{E}[\| \epsilon_n \|_2^2] < \infty, \mathbb{E}[\| \eta_n \|_2^2] < \infty \) so that \( \Sigma = \mathbb{E}\epsilon_n \epsilon_n^\top, \Xi = \mathbb{E}\eta_n \eta_n^\top \) exists. Lemma 1 shows that, if density of \( \epsilon_n, f_x \) exists, then density of \( X_n, n = \ldots, -1, 0, 1, \ldots \) should exist.

**Lemma 1.** Suppose model (6) with \( X_n \) being stationary and suppose distribution of \( \epsilon_n \) being absolutely continuous with respect to Lebesgue measure, then \( X_n \) being absolutely continuous with respect to Lebesgue measure. In particular, density of \( X_n \) exists.

**Proof.** Suppose Borel set \( B \subset \mathbb{R}^d \) satisfying \( m(B) = 0 \) with \( m(.) \) being Lebesgue measure in \( \mathbb{R}^d \), then for any point \( x \in \mathbb{R}^d \), \( m(\{y + x \mid y \in B\}) = 0 \), correspondingly

\[
\text{Prob}(X_{n+1} \in B) = \mathbb{E}\text{Prob}(X_{n+1} \in B \mid X_n) = \mathbb{E} \int_{B-AX_n} f_x \, dm = 0
\]

Thus, distribution of \( X_{n+1} \) is absolutely continuous with respect to Lebesgue measure. From Lebesgue-Radon-Nikodym theorem, we know that density of \( X_{n+1}, f_X \) exists.

With data \( Y_k, k = 1, 2, \ldots, n \) generated by model (6), we can gather consistent estimator of coefficient
matrix $A$ and covariance matrix $\Sigma$ and $\Xi$ according to [13], the estimator is given by

$$\hat{A} = \left(\sum_{k=3}^{n} Y_k Y_k^T \right) \left(\sum_{k=3}^{n} Y_{k-1} Y_{k-2}^T \right)^+$$

and

$$\hat{B}_i = \frac{1}{n} \sum_{k=3}^{n} (Y_k - \hat{A} Y_{k-1})(Y_k - \hat{A} Y_{k-1})^T, \quad i = 1, 2$$

$$\hat{\Xi} = \frac{1}{2} \left( \hat{B}_1 + \hat{A}^T (\hat{B}_1 - \hat{B}_2) \hat{A}^T \right), \quad \hat{\Sigma} = \hat{B}_1 - \hat{\Xi} - \hat{\Xi}^T$$

Here $+$ means pseudo-inverse (see [18] for details). For a random variable $X$, a random vector $Y$ and a constant $\alpha \in (0, 1)$, coincide with (1.2) in [19] and exercise II 1.19 in [20], we define $1 - \alpha$ conditional quantile as

$$c_{1-\alpha}(X|Y = y) = \inf \{x|\text{Prob}(X > x|Y = y) \leq \alpha\}$$

For a function $f$, we define its Fourier transformation and inverse Fourier transformation as

$$\mathcal{F}f(x) = \int_{\mathbb{R}^d} \exp(it^T x)f(t)dt, \quad \mathcal{F}^{-1}f(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x)f(t)dt$$

For function $f, g : \mathbb{R}^d \to \mathbb{C}$, we may define convolution

$$f \ast g(x) = \int_{\mathbb{R}^d} f(y)g(x - y)dy$$

Suppose random variables $X_i, Y_i, i = 1, 2, \ldots, n$ satisfying model [6] and if we suppose density of $\eta_i, f_n$ is known, then with chosen kernel $K$ and bandwidth $h = b(n) = (h_1, \ldots, h_d)$ and $s = s(n) = (s_1, \ldots, s_d)$, the deconvolution density estimator of random variable $X_i$ for given point $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$, $\hat{f}_X(x)$ is defined as (based on [10])

$$\hat{f}_X(x) = \frac{1}{n} \sum_{k=3}^{n} L_n(x - Y_k), \quad L_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x)\mathcal{F}K_n(t)dt$$

$$\hat{f}_X(x) = \max(\Re \hat{f}_X(x_1, \ldots, x_d), 0) \prod_{j=1}^{d} 1_{|x_j| \leq s_j} \int_{x_1}^{x_1} \ldots \int_{x_d}^{x_d} \max(\Re \hat{f}_X(x_1, \ldots, x_d), 0)dx_1 \ldots dx_d = 0$$

with convention $\hat{f}_X(x) = 0$ if $\int_{x_1}^{x_1} \ldots \int_{x_d}^{x_d} \max(\Re \hat{f}_X(x_1, \ldots, x_d), 0)dx_1 \ldots dx_d = 0$

Definition of $K_n$ see [19] and $\Re$ means real part of a complex number. From discussion in [11] residues in model [6] is $\xi_n = Y_{n+1} - AY_n = \epsilon_{n+1} + \eta_{n+1} - A\eta_n \Rightarrow \mathcal{F}f_\epsilon(x) = \mathcal{F}f_\epsilon(x)\mathcal{F}f_\ell(-A^T x)$. Based on this observation we can derive the deconvolution density estimator of innovations $\epsilon_i$. Notice that we do not know the coefficient matrix $A$, so we use predictive roots to create estimator rather than residues.

Definition 2. Suppose notations and conditions in definition [7] we define deconvolution density estimator
A given vector norm in $\mathbb{R}^d$

$\|\cdot\|_2$ Euclidean norm in $\mathbb{R}^d$ and corresponding operator norm for matrix in $\mathbb{R}^{d \times d}$

$\|\cdot\|_{L_p}$ $L_p$ norm for functions defined in $\mathbb{R}^d$ with Lebesgue measure

$\|\cdot\|_{R_p}$ $L_p$ norm for random variable defined in probability space $\Omega$

$c_{1-\alpha}(T|S=s)$ $1-\alpha$ quantile of conditional distribution $\text{Prob}(T|S=s)$

$\mathcal{F}f(x)$, $\mathcal{F}^{-1}f(x)$, $f \ast g$ Fourier transformation, inverse Fourier transformation and convolution

$f_T(x)$ Density of random variable $T$ at point $x \in \mathbb{R}^d$

$f_{T|S}(x|y)$ Conditional density of random variable $T$ given $S$ at point $(T,S) = (x,y)$

$\hat{f}_T(x)$ Deconvolution density estimator of random variable $T$

$\hat{f}_{T|S}(x|y), \hat{c}_{1-\alpha}(T|S=s)$ Estimated conditional density and $1-\alpha$ conditional quantile

$\hat{A}, \hat{\Sigma}, \hat{\Xi}$ Estimator for coefficient matrix $A$, covariance matrix $\Sigma$, $\Xi$ in (6)

$\epsilon_n$, $\eta_n$, $\xi_{n-1}$, $\zeta_{n-1}$ $n$th innovation, error, residue and predictive root defined in (6) and (2)

$\hat{f}_i(x)$ for innovations $\epsilon_i$ as

$$
\hat{f}_i(x) = \frac{1}{n-1} \sum_{k=1}^{n-1} R_k(x - \xi_k), \quad R_k(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x) \frac{\mathcal{F}K_h(t)}{\mathcal{F}f_\eta(t)\mathcal{F}f_\xi(-A^T t)} dt \\
\hat{f}_i(x) = \max(\Re \hat{f}_i(x_1, \ldots, x_d), 0) \prod_{j=1}^d \mathbf{1}_{|x_j| \leq s_j} \max(\Re \hat{f}_i(x_1, \ldots, x_d), 0) dx_1 \ldots dx_d
$$

(14)

with convention $\hat{f}_i(x) = 0$ if denominator is 0

The reason why we truncate the standard deconvolution density estimator $\hat{f}_X$, $\hat{f}_i$ is to make sure the estimated densities are really density function (that is, non-negative and integration equals 1).

To make a summary, we list all frequently used notations in table 1, other symbols will be defined when being used.

We then introduce several necessary assumptions for consistency of bootstrap algorithm. According to [13], in order to acquire consistent estimator of coefficient matrix $A$, one have to assume that coefficient matrix $A$ in (6) being invertible and covariance matrix $\Sigma$ and $\Xi$ exists, with $\Sigma$ being invertible. Combine with requirements for consistent deconvolution given in [16], we give the assumptions as follow:

A1) Suppose data $Y_1, \ldots, Y_n$ are generated through model (6) with invertible coefficient matrix $A$, $\|A\|_2 < 1$, $\Sigma$, $\Xi$ exists and $\Sigma$ is invertible.

A2) Suppose distribution of innovation $\epsilon_i$ and error $\eta_i$, $i = \ldots, -1, 0, 1, \ldots$ are absolutely continuous with respect to Lebesgue measure, and density of $\epsilon_i, \eta_i$ respectively satisfies:

Density of random variable $\eta_i$ is known and there exists a series of constants $\alpha_j, \rho_j \geq 0$, $\beta_j \in \mathbb{R}$ ($\beta_j > 0$ if $\rho_j = 0$), $j = 1, 2, \ldots, d$ and constant $c$ such that $\forall x = (x_1, \ldots, x_d) \in \mathbb{R}^d$,

$$
|\mathcal{F}f_\eta(x_1, \ldots, x_d)| \geq c \prod_{j=1}^d (x_j^2 + 1)^{-\beta_j/2} \exp(-\alpha_j |x_j|^{\rho_j})
$$

(15)
for some constant \( c \), and \( \|f_n(x)\|_{L^\infty} < \infty \).

Density of random variable \( \epsilon_i \) is not known, but there exists constants \( a_j, r_j \geq 0, b_j \in \mathbb{R} \) if \( r_j > 0 \) and \( b_j > 1/2 \) if \( r_j = 0 \) for \( j = 1, 2, \ldots, d \), such that

\[
\sum_{j=1}^{d} \int_{\mathbb{R}^d} |\mathcal{F} f_i(x_1, \ldots, x_d)|^2 (1 + x_j^2)^{\nu} \exp(2a_j |t_j|^\nu) dt_1 \ldots dt_d < \infty \tag{16}
\]

A3) Density \( f_x, f_\eta \) satisfies

\[
\sum_{i=1}^{d} \int_{\mathbb{R}^d} x_i^4 f_i(x_1, \ldots, x_d) dx < \infty, \quad \sum_{i=1}^{d} \int_{\mathbb{R}^d} x_i^4 f_\eta(x_1, \ldots, x_d) dx < \infty \tag{17}
\]

Combine this condition with Cauchy inequality we can show that covariance matrix \( \Sigma \) and \( \Xi \) exist and 
\( \mathbb{E}\|\epsilon_i\|_2^2 < \infty, \quad \mathbb{E}\|\eta_i\|_2^2 < \infty \) for any \( i = \ldots, -1, 0, 1, \ldots \).

A4) Chosen \( \alpha \in (0, 1) \) satisfies for any given \( \delta > 0 \) and \( y \in \{y|f_Y(y) > 0\} \),

\[
\int_{x_1-\alpha(||x_n+1-A\eta_n|||Y_n=y)<|x_1|<x_1-\alpha(||x_n+1-A\eta_n|||Y_n=y)+\delta} f_{\alpha n+1-A\eta_n}|Y_n=x|dy > 0 \tag{18}
\]

K1) The used kernel satisfies, for any \( x = (x_1, \ldots, x_d) \in \mathbb{R}^d \) and \( h = (h_1, \ldots, h_d) \), \( h_i > 0 \forall i \),

\[
K(x_1, \ldots, x_d) = \prod_{j=1}^{d} G(x_j), \quad K_h(x_1, \ldots, x_d) = \prod_{j=1}^{d} \frac{1}{h_j} G\left(\frac{x_j}{h_j}\right) \tag{19}
\]

with \( G \) being one dimensional kernel satisfying \( G \in L_2(\mathbb{R}) \) with respect to Lebesgue measure. \( \mathcal{F}G \) is continuous and compactly supported in \([-a, a]\) with \( a > 1, \mathcal{F}G = 1 \) on \([-1, 1]\).

K2) The used bandwidth \( h = (h_1, \ldots, h_d) \) and \( s = (s_1, \ldots, s_d) \) satisfy \( s_i \to \infty \) and

\[
h_j^{2\beta_j} \exp(-2a_j h_j^{-\nu j}) \times \prod_{j=1}^{d} s_j \to 0
\]

\[
\frac{1}{n} \prod_{k=1}^{d} h_k^{-2\beta_k-1/2+\nu k/2} \exp(2\alpha_k d^{\nu k} h_k^{\rho k}) \times \left( \sum_{k=1}^{d} \frac{1}{h_k^{\rho k}} \right)^{\gamma_1} \exp \left( 4 \sum_{j=1}^{d} \alpha_j a_j^{\nu j} \left( \sum_{k=1}^{d} \frac{1}{h_k^{\rho k}} \right)^{\rho j/2} \right) \prod_{j=1}^{d} s_j \to 0 \tag{20}
\]

Here \( \gamma_1 = 2 \sum_{j=1}^{d} \beta_j + d/4 - \sum_{j=1}^{d} \rho_j/4 \) for \( i = 1, 2, \ldots, d \).

Coincide with [21], in assumption A2), when \( \rho_j = 0 \) for \( \forall j = 1, 2, \ldots, d \), the error distribution \( f_\eta \) is called ordinary smooth, when \( \exists j \) such that \( \rho_j > 0 \), the error distribution is called super smooth. Smoothness of error distribution influences the bandwidth chosen in definition [1] and the convergence rate of density estimator. When error distribution is super smooth, the convergence rate is of \( \log(n) \), which severely influences estimation of density and performance of bootstrap algorithm. Condition A4) coincides with strictly increasing assumption used in theorem 1.2.1, [19]. It is relatively abstract and difficult to justify, so we provide a stricter condition that is easy to check in corollary [1].

According to Plancherel theorem (theorem 8.29 in [17] and chapter 2.2.4 in [22]), Fourier transformation of
$L^2$ function exists and inversion formula holds almost surely. One of the kernel $G$ satisfying conditions K1) is

$$G(x) = \frac{1}{\pi x^2} (\cos(x) - \cos(2x)),$$

$$FG(x) =
\begin{cases} 
1 & x \in [-1, 1] \\
2 - x & x \in [1, 2] \\
x + 2 & x \in [-2, -1] \\
0 & \text{Otherwise}
\end{cases}$$

which is actually $L^1 \cap L^2$ function.

## 2 Prediction with dependent data

Generally speaking, prediction problem with dependent data involves generating predictor of next observation and estimating conditional distribution and quantile of predictive root (definition see [1]). If these two things are estimated well, then we can construct 'confidence region' for next observation (always called prediction region), to make sure that next observation belongs to this region with desired probability.

As mentioned in 1.1, conditional density of $\eta_n$ given observation $Y_n$ at point $(\eta_n, Y_n) = (x, y_n)$ is $f_{\eta_n\mid Y_n}(x|y_n) = \frac{f_{\eta,n}(x)}{f_{\eta,n}(y_n)}$, which depends on previous observation. Combine with (2), conditional distribution of residues will also depend on previous observation. The last observation is a random variable, so it is not reasonable to assume conditional quantile $c_{1-\alpha}(X_{n+1}\mid Y_n)$ as a constant. Worse still, empirical distribution of residues $Y_{k+1} - AY_k$, $k = 1, 2, ..., n-1$ estimates marginal distribution of them rather than conditional one, so residue-based bootstrap is not suitable. Because conditional quantile is a non-degenerated random variable, definition and asymptotically valid condition frequently used in prediction (for example, definition 2.3 in [1]) should be clarified. In this chapter, we will first introduce definition of consistent prediction quantile and then provide an asymptotic valid condition for consistency.

**Definition 3** (Consistent prediction quantile). Suppose we have observations $Y_k$, $k = 1, 2, ..., n$ from a stationary model, for a chosen norm $\| \|$ in $\mathbb{R}^d$ and given $\alpha \in (0, 1)$, new observation $X_{n+1}$ and predictor $\tilde{X}_{n+1}$, if estimated and true quantile function satisfies

$$\left| c_{1-\alpha}(\|X_{n+1} - \tilde{X}_{n+1}\| \mid Y_1, ..., Y_n) - c_{1-\alpha}(\|X_{n+1} - \tilde{X}_{n+1}\| \mid Y_1, ..., Y_n) \right| \overset{\text{Pr}}{\longrightarrow} 0$$

then we say estimated $1 - \alpha$ quantile being consistent. Here $\overset{\text{Pr}}{\longrightarrow}$ means converging in probability as sample size $n \to \infty$.

**Remark 1.** Quantiles in (22) are estimated and true quantile of conditional distribution $\|X_{n+1} - \tilde{X}_{n+1}\|$ given previous observation $Y_1, ..., Y_n$. In state space model (6), definition 3 means that estimated and true quantile of $\|X_{n+1} - \tilde{X}_{n+1}\|\mid Y_n$ should not be far away from each other for the majority of possible gathered $Y_n$.

**Remark 2.** If $X_{n+1} - \tilde{X}_{n+1}$ is asymptotically independent of previous observations (like autoregressive models), then $c_{1-\alpha}(\|X_{n+1} - \tilde{X}_{n+1}\| \mid Y_1, ..., Y_n)$ is asymptotically a constant and definition 3 coincides with frequently used asymptotic valid condition for prediction problem introduced in [7].
Because we do not want to directly deal with quantiles, we provide a sufficient condition for consistency based on conditional density under model (6), which is easier to deal with. First we give a lemma that describes continuity of conditional quantile.

**Lemma 2.** Suppose data $Y_1, Y_2, \ldots$ are gathered from model (6), $\alpha \in (0, 1)$ is chosen so that conditional density $f_{X_{n+1}-AY_n|Y_n} (x|y)$ satisfies condition A4). Suppose $f_{X_{n+1}-AY_n|Y_n} (x|y)$ is continuous in $\mathbb{R}^d \times \{y|f_Y(y) > 0\}$ and $f_Y(y)$ is continuous on $\mathbb{R}^d$, then for any $y$ such that $f_Y(y) > 0$, conditional quantile $c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n = y)$ is continuous at $Y_n = y$.

Continuity of $f_Y$ implies that set $\{y|f_Y(y) > 0\}$ is open, so continuity of conditional quantile is well-defined. The next lemma involves uniform convergence result in compact sets. According to [5], $X_{n+1} - AY_n = \epsilon_{n+1} - A\eta_n$, and we will use $f_{X_{n+1}-AY_n|Y_n}$ to represent conditional density in lemma 2 in the following theorems.

**Lemma 3.** For a given compact set $M \subset \{y|f_Y(y) > 0\}$, suppose constant $\alpha$ and conditional density function $f_{X_{n+1}-AY_n|Y_n} (x|y)$ satisfy conditions in lemma 2, in addition suppose a sequence of functions $\{g_m(x|y)\}_{m=1}^\infty$ satisfying

1) for any given $y$ such that $f_Y(y) > 0$, $g_m(x|y) \geq 0$ for any $x \in \mathbb{R}^d$ and $\int_{\mathbb{R}^d} g_m(x|y) dx = 1$

2) There exists a constant $1 \leq p \leq \infty$ such that as $m \to \infty$,

$$
\sup_{y \in M} \|g_m(.|y) - f_{X_{n+1}-AY_n|Y_n}(.|y)\|_{L_p} \to 0 \tag{23}
$$

Define $d_{m,1-\alpha}(y)$ being $1 - \alpha$ quantile function of function $g_m(x|y)$, we have

$$
\sup_{y \in M} |d_{m,1-\alpha}(y) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n = y)| \to 0 \tag{24}
$$

as $m \to \infty$. $L_p$ norm in (23) is taken with respect to $x$.

With lemma 2 and 3, we can describe the key theorem in this chapter, which connects convergence of conditional quantile with convergence of conditional density.

**Theorem 1.** Suppose conditional density of observed data, $f_{X_{n+1}-AY_n|Y_n}(x|y)$ satisfies conditions in lemma 2 and suppose for any given compact set $K \subset \{y|f_Y(y) > 0\}$, the estimated conditional density $\hat{f}(x|y)$ satisfies

1) for any given compact set $K \subset \{y|f_Y(y) > 0\}$, $\text{Prob}(\exists x \in \mathbb{R}^d, y \in K, \hat{f}(x|y) < 0) \to 0$ and $\text{Prob}(\exists y \in K, \int_{\mathbb{R}^d} \hat{f}(x|y) dx \neq 1) \to 0$ as $n \to \infty$. If the sets defined above are not measurable, $\text{Prob}$ is defined as outer measures.

2) 

$$
\sup_{y \in K} \|\hat{f}(.|y) - f_{X_{n+1}-AY_n|Y_n}(.|y)\|_{L_p} \xrightarrow{n \to \infty} 0 \tag{25}
$$

then we have, for any given $\alpha \in (0, 1)$, conditional quantile function of estimated density $\hat{f}(x|y)$, $\hat{c}_{1-\alpha}(y)$ satisfies

$$
|\hat{c}_{1-\alpha}(Y_n) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)| \xrightarrow{n \to \infty} 0 \tag{26}
$$

Here $\hat{c}_{1-\alpha}, c_{1-\alpha}$ are two functions of last observation $Y_n$ and therefore they are non-degenerated random variables.
Figure 1: Illustration for condition A4). Blue and orange line respectively represents conditional density with different $y$. In this example, 0.98 quantile of blue density is less than 1 but 0.98 quantile of orange density is larger than 1.5.

Remark 3. Condition A4) is a counterpart of (43), which automatically holds according to (10). The reason why we propose this condition is that intervals with 0 density make quantile function discontinuous. For example, in figure 1 density with blue line is 0 outside interval $[0, 1]$, so it is possible for the orange density to be close to blue density in $L_p$ norm but has a region with positive probability which is far away from $[0, 1]$.

Remark 4. Theorem 1 tells us that we can separate possible values the last observation may acquire into two regions and if we can make sure that estimators work well uniformly in one of the region and the probability of the other region is asymptotically negligible, then the estimated prediction region should work well.

3 Bootstrap 1 step prediction region for state space model

In this section, we propose a bootstrap algorithm (algorithm 1) for constructing 1 step prediction region satisfying 3 and provide theoretical proof of consistency. Theorem 3 shows that estimators defined in definition 1 and 2 are consistent with true densities in $L_1$ norm. Lemma 4 gives conditional density of prediction root $X_{n+1} - AY_n$ conditioning on $Y_n$.

Lemma 4. Suppose $X_{n+1}$ and $Y_n$ are generated from model (6), then conditional density of prediction root $X_{n+1} - AY_n$ on $Y_n$ at point $(X_{n+1} - AY_n, Y_n) = (x, y)$ satisfies

$$f_{X_{n+1} - AY_n|Y_n}(x|y) = \frac{\int_{\mathbb{R}^d} f_x(t) f_\eta(A^{-1} t - A^{-1} x) f_X(y + A^{-1} x - A^{-1} t) dt}{|\det(A)| \int_{\mathbb{R}^d} f_x(t) f_\eta(y - t) dt}$$

(27)

with the convention if $\int_{\mathbb{R}^d} f_x(t) f_\eta(y - t) dt = 0$, then $f_{X_{n+1} - AY_n|Y_n}(x|y) = 0$.

Proof. Notice that $X_{n+1} - AY_n = \epsilon_{n+1} - A\eta_n$ and $Y_n = X_n + \eta_n$, so joint density of $\epsilon_{n+1}, \eta_n, X_n$ is $f_\epsilon f_\eta f_X$ and marginal density of $Y_n$ is

$$f_Y(y) = \int_{\mathbb{R}^d} f_x(t) f_\eta(y - t) dt$$

(28)

which exists from Young’s inequality. According to change of variable theorem (theorem 2.44 in [17]), joint
density of $X_{n+1} - AY_n = c_{n+1} - A\eta_n, Y_n$ is

$$f_{n+1 - A\eta_n, Y_n}(x, y) = \frac{1}{|\det(A)|} \int_{\mathbb{R}^d} f_\epsilon(t) f_\gamma(A^{-1}t - A^{-1}x)f_X(y - A^{-1}t + A^{-1}x)$$

Thus, when (28) is not 0, we have (31) holds for points $x$ such that $1/n < 0$

Proof. By change function value of points in a 0 measure set, we may assume $f_\epsilon > 0, \forall x \in \mathbb{R}^d$. Change of variable theorem shows that $\frac{1}{|\det(A)|} \int_{\mathbb{R}^d} f_\epsilon(t) f_\gamma(A^{-1}t - A^{-1}x)f_X(y + A^{-1}x - A^{-1}t)dt = \int_{\mathbb{R}^d} f_\epsilon(Az + x)f_\gamma(fz)(y - z)dz$. From continuity of measure and theorem 2.44 in [17], we have for any given $x \in \mathbb{R}^d$ and $y$ such that $\int_{\mathbb{R}^d} f_\epsilon(yz)(y - z)dz > 0$,

$$0 < m(\{z| f_\gamma(fz)(y - z) > 1/n\} \cap \{z| f_\epsilon(Az + x) > 1/n\})$$

$$\Rightarrow \int_{\mathbb{R}^d} f_\epsilon(Az + x)f_\gamma(fz)(y - z)dz \geq \frac{1}{n^2} m(\{z| f_\gamma(fz)(y - z) > 1/n\} \cap \{z| f_\epsilon(Az + x) > 1/n\}) > 0$$

for sufficiently large $n$ such that $m(\{z| f_\gamma(fz)(y - z) > 1/n\} \cap \{z| f_\epsilon(Az + x) > 1/n\}) > 0$. Especially we have (31) holds for points $x$ such that $c_{1-\alpha}(\|c_{n+1} - A\eta_n\|Y_n = y) < \|x\| < c_{1-\alpha}(\|c_{n+1} - A\eta_n\|Y_n = y) + \delta$ and condition A4) holds.

Corollary 1. Suppose conditions A1), A2) and suppose density $f_\epsilon$ being positive almost surely with respectively to Lebesgue measure in $\mathbb{R}^d$, then condition A4) is satisfied with any $\alpha \in (0, 1)$.

Algorithm 1 (Bootstrap algorithm for prediction interval).

Input: Data $Y_i, i = 1, 2, ..., n$ satisfying model (6), density of errors $f_\epsilon$, re-sample times $m$ satisfying $1/m \ll \alpha$, a large integer $N$ and bandwidth $h(n) = (h_1, ..., h_d)$, $s(n) = (s_1, ..., s_d)$ satisfying K2).

1) Estimate coefficient matrix $\hat{A}$ and covariance matrix $\hat{\Sigma}$ based on [8] and [9].

2) Randomly choose $X_0^* \in \mathbb{R}^d$ (for example, let $X_0^* = \mathbf{0}$), for $i = 1, 2, ..., N + n$, generate $\epsilon_i^*$ from i.i.d. multivariate standard normal distribution $N(0, I)$, here $I$ is $d \times d$ identity matrix. We let $X_{i+1}^* = \hat{A}X_i^* + \sqrt{\hat{\Sigma}}\epsilon_{i+1}$ for $i = 0, 1, ..., N + n - 1$. Generate $\eta_j^*, j = 1, 2, ..., n$ independently from known density $f_\gamma$, let $Y_j^* = X_{j+n}^* + \eta_j^*$ for $j = 1, 2, ..., n$. Definition of matrix square root see [23].

3) Recalculate coefficient matrix $\hat{A}^*$ from [8] based on bootstrap data $Y_j^*$, $j = 1, 2, ..., n$, and the predictor for $X_{n+1}^*$ is given by $\hat{A}Y_n$.

4) Generate $\epsilon_{n+1}^*$ by density $\hat{f}_\epsilon$, generate $\gamma_{n+1}^*$ by density $\hat{f}_{\gamma}(Y_n | \epsilon_n) = \frac{f_X(Y_n - \epsilon_n)f_\gamma(\epsilon_n)}{\int_{\mathbb{R}^d} f_X(Y_n - \epsilon_n)f_\gamma(\epsilon_n)d\epsilon}$

5) Calculate $X_{n+1}^* = \hat{A}Y_n + \epsilon_{n+1}^* - A\gamma_{n+1}^*$, calculate $\tau_\star = \|A_{n+1} - \hat{A}Y_n\|

6) Repeat 2) to 5) for $m$ times and gather $\tau_j^\star = \|X_{n+1,j} - \hat{A}Y_n\|, j = 1, 2, ..., m$ such that $\tau_1^\star \leq \tau_2^\star \leq \ldots \leq \tau_m^\star$, we acquire $1 - \alpha$ sample quantile $c_{1-\alpha}^\star = \min_{j=1,2,\ldots,m} \{\tau_j^\star \frac{m-1}{m} \leq \alpha\}$. 10
7) Consistent prediction region is given by \( \{ x \in \mathbb{R}^d | \| x - \hat{A}Y_n \| \leq c_{1-a} \} \)

The reason why we choose a large integer \( N \) and resample \( N + n \) times is to avoid the influence of initial value \( X_0^* \). Because the influence has exponential decay, in practice choosing \( N \) as several hundred is sufficient.

In theoretical justification, first thing is to show that estimated density \( \hat{f}_s \) and \( \hat{f}_X \) are really densities functions (that is, non-negative and integration being 1) and converge to target densities. We show it in theorem 2 and corollary.

**Theorem 2.** Suppose \( n \) observed data \( Z_i, i = 1, 2, ..., n \) satisfies \( Z_i = S_i + T_i \) with random vectors \( S_i, T_i \in \mathbb{R}^d, i = 1, 2, ..., n \) being stationary and \( S_i \) being independent with \( T_j, j = 1, 2, ..., n \) for \( \forall i = 1, 2, ..., n \) (but \( S_i, i = 1, 2, ..., n \) and \( T_j, j = 1, 2, ..., n \) may depend on each other respectively). Suppose \( Z_i \) being \( \alpha \)-mixing with mixing coefficients \( \alpha_Z(k), k \in \mathbb{Z}^+ \) such that

\[
\sum_{k=1}^{\infty} \alpha_Z(k) < \infty \tag{32}
\]

In addition, suppose distribution of \( S_i \) and \( T_i \) are absolutely continuous with respect to Lebesgue measure and densities \( f_S, f_T \) satisfy for any \( x = (x_1, ..., x_d) \),

\[
\sum_{i=1}^{d} \int_{\mathbb{R}^d} |\mathcal{F}f_S(t_1, ..., t_d)|^2 (1 + |t|^2)^{\frac{d}{2}} \exp(2a_i|t|^\rho) dt_1 ... dt_d < \infty
\]

\[
|\mathcal{F}f_T| \geq \epsilon \prod_{j=1}^{d} (x_j^2 + 1)^{-\beta_j/2} \exp(-\alpha_j|x_j|^\rho) \tag{33}
\]

with constants \( b_i, a_i, \rho_i, \alpha_i, i = 1, 2, ..., d \), kernel \( K_h \) satisfy A2), K1 and constant \( c > 0 \). For any \( x = (x_1, ..., x_d) \in \mathbb{R}^d \), define

\[
\hat{f}_S(x) = \frac{1}{n} \sum_{h=1}^{n} L_h(x - Z_h), \quad L_h(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x) \mathcal{F}K_h(t) \frac{d}{dt} \mathcal{F}f_T(t) dt
\]

\[
\hat{f}_S(x_1, ..., x_d) = \max(\text{Re}f_s(x_1, ..., x_d), 0) \prod_{i=1}^{d} 1_{|\rho_i| \leq \alpha_i} \int_{x_1}^{x_1} ... \int_{x_d}^{x_d} \max(\text{Re}f_s(x_1, ..., x_d), 0) dx_1 ... dx_d \tag{34}
\]

with convention \( \hat{f}_S(x_1, ..., x_d) = 0 \) if \( \int_{x_1}^{x_1} ... \int_{x_d}^{x_d} \max(\text{Re}f_s(x_1, ..., x_d), 0) dx_1 ... dx_d = 0 \), then we have

1) There exists a constant \( C \) independent of sample size \( n \) and chosen bandwidth \( h = h(n) = (h_1, ..., h_d) \) such that

\[
\mathbb{E} \| \hat{f}_S - f_S \|_{L_2}^2 \leq C \left( \frac{1}{n} \prod_{j=1}^{d} h_j^{-2\beta_j-1+\rho_j} \exp(2\alpha_j q^{\rho_j} h_j^{-\rho_j}) + \sum_{j=1}^{d} h_j^{2\rho_j} \exp(-2\alpha_j h_j^{-\rho_j}) \right) \tag{35}
\]

2) If bandwidth \( h = h(n) = (h_1, ..., h_d) \) and \( s = s(n) = (s_1, ..., s_d) \) satisfies for \( \forall i = 1, 2, ..., d \), \( s_i \to \infty \) and

\[
\frac{1}{n} \prod_{j=1}^{d} s_j h_j^{-2\beta_j-1+\rho_j} \exp(2\alpha_j q^{\rho_j} h_j^{-\rho_j}) + \sum_{j=1}^{d} h_j^{2\rho_j} s_1 ... s_d \exp(-2\alpha_j h_j^{-\rho_j}) \to 0 \tag{36}
\]

as \( n \to \infty \), then

\[
\| \hat{f}_S - f_S \|_{L_1} \xrightarrow{P_{\text{prob}}} 0 \tag{37}
\]
Lemma 5. Suppose data $Y_i, i = 1, 2, \ldots, n$ are generated by model $\rho$ and condition \text{A3}) is satisfied, then there exists a constant $C$ such that for any $n \geq 3$ and sufficiently small $s$,

$$\text{Prob}(\|\hat{A} - A\|_2 \geq s) \leq \frac{C}{(n - 2)s^2} \quad (38)$$

With the convergence rate of parameter matrix, we can prove the consistency of estimated density of innovations as well as quantiles. Theorem 3 is our main result, which directly shows that quantile generated by bootstrap algorithm satisfies definition 5.

**Theorem 3.** Suppose gathered data $Y_i, i = 1, 2, \ldots, n$ are stationary, distributions of $\epsilon_i, \eta_i, i = \ldots, -1, 0, 1, \ldots, \alpha \in (0, 1)$ and kernel $K_1, \ldots, K_2$, bandwidth $h(n), s(n)$ satisfy conditions A1) to A4), K1), K2), then we have

1) Estimator defined in definition 7 and 3 satisfy

$$\|\hat{f} - f \|_{L_1} \xrightarrow{\text{Prob}} 0, \|\hat{f}_X - f_X\|_{L_1} \xrightarrow{\text{Prob}} 0, \|\hat{f} \ast f_0(\cdot) - f_0(\cdot)\|_{L_{\infty}} \xrightarrow{\text{Prob}} 0 \quad (39)$$

2) Conditional density $f_{X_{n+1}^{*}} - \hat{A}Y_n | Y_1, \ldots, Y_n$ of bootstrap random variable $X_{n+1}^{*} - \hat{A}Y_n | Y_1, \ldots, Y_n$ defined in bootstrap algorithm 4) satisfies conditions 1) and 2) in lemma 2.

3) Conditional quantile of random variable $\|X_{n+1}^{*} - \hat{A}Y_n\| | Y_1, \ldots, Y_n$ and $\|X_{n+1} - \hat{A}Y_n\| | Y_n$ satisfy

$$|c_{1-\alpha}(\|X_{n+1} - \hat{A}Y_n\| | Y_1, \ldots, Y_n) - \hat{c}_{1-\alpha}(\|X_{n+1}^{*} - \hat{A}Y_n\| | Y_1, \ldots, Y_n)| \xrightarrow{\text{Prob}} 0 \quad (40)$$

In bootstrap world, we can independently generate infinity replicates of random variables whose distribution coincides with $\|X_{n+1}^{*} - \hat{A}Y_n\| | Y_1, \ldots, Y_n$, and remark 5 tells us that quantile $c_{1-\alpha}$ generated by bootstrap algorithm converges to underlying quantile of distribution of $\|X_{n+1} - \hat{A}Y_n\| | Y_1, \ldots, Y_n$ at continuous points of $h(\alpha) = \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A}Y_n\| | Y_1, \ldots, Y_n)$.

**Remark 5.** According to Glivenko-Cantelli theorem (theorem 19.1 in [23]), suppose the independent bootstrap replicates in algorithm 7 for $Z = \|X_{n+1}^{*} - \hat{A}Y_n\| \leq x | Y_1, \ldots, Y_n$ are $Z_i^{*}; i = 1, 2, \ldots, m$, we have

$$\sup_{x \geq 0} \frac{1}{m} \sum_{j=1}^{m} 1_{Z_j^{*} \leq x} - \text{Prob}(\|X_{n+1}^{*} - \hat{A}Y_n\| \leq x | Y_1, \ldots, Y_n) \xrightarrow{a.s.} 0 \quad (41)$$

Define $F(x) = \text{Prob}(\|X_{n+1}^{*} - \hat{A}Y_n\| \leq x | Y_1, \ldots, Y_n)$ and $h(\alpha) = \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A}Y_n\| | Y_n)$. If $F$ is continuous
Table 2: Numerical result of bootstrap algorithm 1 for model (6). (R) means that prediction region is generated by residue-based bootstrap.

| shape & scale | sample size | False rate | Estimated coefficient | Last observation $Y_n & X_n$ | 95% quantile |
|--------------|-------------|------------|-----------------------|-----------------------------|--------------|
| 0.2, 5.0     | 600000     | 0.075      | 0.7006                | 1.403, 1.403                | 3.242        |
| 0.2, 5.0     | 100000     | 0.037      | 0.6925                | 1.259, 1.188                | 3.206        |
| 0.2, 5.0     | 200000     | 0.040      | 0.6909                | -1.596, -1.282              | 3.278        |
| 0.05, 5.0(R) | 200000     | 0.009      | 0.6909                | -1.596, -1.282              | 4.700        |
| 0.05, 5.0     | 100000     | 0.022      | 0.7098                | 2.674, 2.437                | 3.778        |
| 0.05, 5.0     | 50000      | 0.038      | 0.7023                | -0.621, -0.621              | 3.219        |
| 0.05, 5.0     | 50000      | 0.037      | 0.6998                | 2.394, 2.049                | 3.641        |
| 0.05, 5.0     | 300000     | 0.022      | -0.621                | 2.394, 2.049                | 3.771        |
| 0.05, 5.0     | 600000     | 0.038      | 0.7023                | -1.596, -1.282              | 3.219        |

and strictly increasing at point $h(\alpha)$, then (41) with lemma 1.2.1 in [19] implies that $c_{1-\alpha}$ generated in algorithm 1 satisfies $|c_{1-\alpha}^* - h(\alpha)|_{m \rightarrow \infty} \rightarrow 0$ and $F(c_{1-\alpha}^*)_{m \rightarrow \infty} \rightarrow F(h(\alpha)) = 1 - \alpha$ in bootstrap world, which is asymptotic valid condition defined in [17]. Thus, if we can show that underlying quantile $c_{1-\alpha}(X_{n+1}^* - \hat{A}Y_n), \ldots, Y_n$ $\overset{Prob}{\rightarrow} c_{1-\alpha}(X_{n+1}^* - \hat{A}Y_n)$ in the original probability space, then for sufficiently large number of replicates, $c_{1-\alpha} \approx \hat{c}_{1-\alpha}(X_{n+1}^* - \hat{A}Y_n)$ should be useful for constructing confidence region.

4 Numerical examples

The purpose of this section is to demonstrate the consistency of bootstrap algorithm 1 and illustrate that smoothness of Fourier transformation of errors $\eta_n$, $n = 1, 2, \ldots$ influences performance of bootstrap algorithm 1. In the following examples, we prefer acceptance/rejection method to generate random variable (see algorithm 7.1 in [25]).

Example 1. In this example, we suppose coefficient matrix $A = 0.7$. Following [19], innovation $\epsilon_n$ is generated by standard Laplace distribution whose characteristic function is $\mathcal{F}(x, y) = \frac{1}{2} \frac{1}{1 + x^2}$ with parameter $(a_1, b_1, r_1) = (0, t, 0)$ with $t < 3/2$. Error $\eta_i, i = 1, 2, \ldots, n$ are generated as $\eta_i = r_{1,i} - r_{2,i}$, where $r_{1,1}, r_{1,2}, i = 1, 2$ being independent and satisfy gamma distribution with shape $k$ and scale $\theta$. Fourier transformation of $f_\eta$ is $\mathcal{F}(x) = (1 + \theta^2 x^2)^{-k}$, so $(a_1, \beta_1, \rho_1) = (0, 2k, 0)$. Bandwidth $h(n)$ and $s(n)$ are chosen as $h(n) = n^{-1/5}$, $s(n) = \log(n)$ satisfying condition K2) with chosen shape and scale. The result is demonstrated in table 2. We also run residue-based bootstrap (algorithm 3.2 in [2]) as a comparison. False rate is defined as the percentage of realizations $X_{n+1}$ to be out of prediction region, thus for 95% prediction region false rate should be close to 0.05.

In example 1, table 2 and figure 1 shows that residue-based bootstrap is not consistent for model (6) while algorithm 1 has a better performance. Besides, smoothness of errors directly affect performance of algorithm 1 and finally, when the underlying $X_n$ deviates largely from observation $Y_n$ (like figure 2(c)), finite sample behavior of algorithm 1 will be influenced.

Example 2. Example 2 demonstrates performance of algorithm 1 on two dimension data. In this experiment, we choose distribution of innovations as product measures of standard Laplace distribution. Thus density
Figure 2: Numerical prediction regions for $X_{n+1}$ generated by different algorithms, red lines, green dashed lines, black dot-dashed lines and purple dot lines respectively denote boundary for 91%, 93%, 95%, 97% prediction region, sample size is 200000 for (a) and (b), 600000 for (c), detail parameters see table 2 3rd, 4th and 9th lines.
Table 3: Numerical result of bootstrap algorithm for two dimensional data

| shape & scale | sample size | False rate | Estimated coefficient | Last observation $Y_n$ & $X_n$ | 95% quantile |
|--------------|-------------|------------|-----------------------|-------------------------------|--------------|
| 0.025 & 7.0  | 300000      | 0.0839     | 0.142, 0.265          | (-1.174, 0.886)$^T$          | 3.620        |
|              |             |            | 0.568, 0.348          | &(-1.172, 0.886)$^T$         |              |
| 0.025 & 7.0  | 800000      | 0.0718     | 0.132, 0.271          | (0.592, 0.137)$^T$           | 3.761        |
|              |             |            | 0.543, 0.362          | & (0.592, 0.725)$^T$         |              |
| 0.05 & 7.0   | 300000      | 0.0807     | 0.160, 0.260          | (-1.894, 0.791)$^T$          | 3.732        |
|              |             |            | 0.549, 0.365          | & (-0.962, 0.792)$^T$        |              |
| 0.05 & 7.0   | 800000      | 0.0568     | 0.145, 0.271          | (1.739, 0.740)$^T$           | 3.957        |
|              |             |            | 0.583, 0.333          | & (1.738, 0.750)$^T$         |              |

Figure 3: Numerical prediction region generated by algorithm for $X_{n+1}$, infinity norm is used and meaning of lines is the same as 1 detail information see 2nd line in table 3

Distribution of innovations is $f(x, y) = \frac{1}{2} \exp(-|x| - |y|)$ and $(\alpha_i, b_i, r_i) = (0, t, 0), t < 3/2, i = 1, 2$. Distribution of errors is generated as product measures of distribution in example 2, so its Fourier transformation is given by $(1 + \theta^2 x^2)^{-k} (1 + \theta^2 y^2)^{-k}$, so that $(\alpha_i, \beta_i, \rho_i) = (0, 2k, 0), i = 1, 2$. Coefficient matrix is given by

$$A = \begin{pmatrix} 0.14 & 0.27 \\ 0.56 & 0.35 \end{pmatrix}$$

and bandwidth is chosen as $h(n) = n^{-1/7}$ and $s(n) = \log(n)/3$. Performance of algorithm on these data is demonstrated in table 3 and figure 3. Influence of smoothness of error distribution is not so significant as in example 7.

5 Conclusion

In this paper, we propose a bootstrap algorithm for prediction problem under simplified state space model and prove asymptotic consistency of the algorithm. Besides, we discuss prediction problem under the situation
that innovation depends on previous observations and extend definition in [1] to a broader class of models. We also extend results in [16] to weakly dependent data. Numerical experiments show that smoothness of errors affects convergence rate of algorithm [1] and the algorithm will have a slow convergence rate when $X_n$ deviates largely from last observation $Y_n$. Apart from state space model, results in this paper can be adjusted to stochastic volatility model [26], exogenous time series model [27] and etc.

6 Appendix

Definition set $S_Y = \{ y | f_Y(y) > 0 \}$ and $c_{1-\alpha}(y) = c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n = y)$ for any $y \in S_Y$. Notice that $f_Y$ is continuous, so $S_Y$ is open. For any $y \in S_Y$, there exists an open ball $B(y, r)$ such that $f_Y(x) > 0$ and conditional quantile is well-defined for $\forall x \in B(y, r)$. In finite dimension space $\mathbb{R}^d$, all norms are equivalent (Chapter 5, exercise 6 in [17]), so it suffices to show that for $\forall 0 < \varepsilon < r/2$, there exists an $\xi > 0$ sufficiently small such that conditional quantile is well-defined in $B(y, \xi)$ and $\forall z, \|z - y\| < \xi \Rightarrow |c_{1-\alpha}(y) - c_{1-\alpha}(z)| < \varepsilon$. From definition [10], for any $\delta > 0$ and $y \in \text{supp} f_Y_n$,

$$\int_{c_{1-\alpha}(y) - \delta \leq \|x\| < c_{1-\alpha}(y)} f_{n+1 - A\eta_n | Y_n}(x, y)dx > 0 \quad (43)$$

otherwise there exists $\delta > 0$ such that

$$\text{Prob}(\|X_{n+1} - AY_n\| > c_{1-\alpha}(y) - \delta) = \int_{c_{1-\alpha}(y) - \delta < \|x\|} f_{n+1 - A\eta_n | Y_n}(x, y)dx = \int_{c_{1-\alpha}(y) - \delta < \|x\|} f_{n+1 - A\eta_n | Y_n}(x, y)dx \quad (44)$$

From dominated convergence theorem, we have

$$\int_{c_{1-\alpha}(y) - \delta < \|x\|} f_{n+1 - A\eta_n | Y_n}(x, y)dx = \lim_{\varepsilon \to 0^+} \int_{c_{1-\alpha}(y) - \varepsilon < \|x\|} f_{n+1 - A\eta_n | Y_n}(x, y)dx \leq \alpha \quad (45)$$

and (44) contradicts with definition of $c_{1-\alpha}(y)$.

Notice that conditional density $f_{n+1 - A\eta_n | Y_n}(x | y)$ is continuous, so it is uniformly continuous in compact set $\|\|^{-1}([0, c_{1-\alpha}(y) + 1]) \times \{ z | \|z - y\| < \|x\| \leq r/2 \}$, here $r$ satisfies $B(y, r) \subset S_Y$. There exists a constant $\xi < r/2$ such that

$$\|z - y\| < \xi \Rightarrow \|f_{n+1 - A\eta_n | Y_n}(x | y) - f_{n+1 - A\eta_n}(x | z)\| < \frac{\min(\gamma, \rho)}{m([\|\|^{-1}([0, c_{1-\alpha}(y) + r])))} \quad (46)$$

$$\gamma = \int_{c_{1-\alpha}(y) - \delta < \|x\| < c_{1-\alpha}(y) + \delta} f_{n+1 - A\eta_n | Y_n}(x | y)dx, \quad \rho = \int_{c_{1-\alpha}(y) - \delta < \|x\| < c_{1-\alpha}(y)} f_{n+1 - A\eta_n | Y_n}(x | y)dx$$

for any $x$ such that $\|x\| \leq c_{1-\alpha}(y) + r/2$, here $m(.)$ denotes Lebesgue measure in $\mathbb{R}^d$. We choose this $\xi$, for
any \(z\) such that \(\|y - z\|_2 < \xi\),

\[
\text{Prob}(\|X_{n+1} - A\| > c_{1-\alpha}(y) + \epsilon |Y_n = z) = \int_{\|x\| > c_{1-\alpha}(y) + \epsilon} f_{n+1, A|Y_n}(x|z)dx
\]

\[
\leq \alpha - \int_{c_{1-\alpha}(y) \leq \|x\| < c_{1-\alpha}(y) + \epsilon} f_{n+1, A|Y_n}(x|y)dx + \int_{\|x\| \leq c_{1-\alpha}(y) + \epsilon} f_{n+1, A|Y_n}(x|y)dx - f_{n+1, A|Y_n}(x|z)dx < \alpha \tag{47}
\]

Thus, we have \(c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = z) \leq c_{1-\alpha}(y) + \epsilon\).

On the other hand,

\[
\text{Prob}(\|X_{n+1} - A\| > c_{1-\alpha}(y) - \epsilon |Y_n = z) = \int_{\|x\| > c_{1-\alpha}(y) - \epsilon} f_{n+1, A|Y_n}(x|z)dx
\]

\[
\geq \int_{\|x\| > c_{1-\alpha}(y) - \epsilon} f_{n+1, A|Y_n}(x|y)dx - \int_{\|x\| \leq c_{1-\alpha}(y)} f_{n+1, A|Y_n}(x|y)dx - f_{n+1, A|Y_n}(x|z)dx > \alpha \tag{48}
\]

Thus, we have \(c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = z) \geq c_{1-\alpha}(y) - \epsilon\). Combine with (47) and (48), we prove the result. 

\(\square\)

**proof of lemma 3.** Similar as lemma 2, we denote \(c_{1-\alpha}(y) = c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = y)\) and \(S_Y = \{y | f_Y(y) > 0\}\). According to lemma 2, conditional \(1 - \alpha\) quantile \(c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = y)\) is a continuous function. For any given \(\delta > 0\) and \(\alpha \in (0, 1)\), for \(\forall y \in S_Y\), define function

\[
h_{+, \delta}(y) = \int_{c_{1-\alpha}(y) + \delta < \|x\| < c_{1-\alpha}(y) + \delta} f_{X_{n+1}, A|Y_n}(x|y)dx > 0
\]

\[
h_{-, \delta}(y) = \int_{c_{1-\alpha}(y) - \delta < \|x\| < c_{1-\alpha}(y) - \delta} f_{X_{n+1}, A|Y_n}(x|y)dx > 0 \tag{49}
\]

For any \(y \in S_Y\), \(\forall z \in \mathbb{R}^d\) such that \(\|y - z\|_2\) being sufficiently small so that conditional quantile \(c_{1-\alpha}(z)\) being well defined and any given \(\epsilon > 0\), without loss of generality, suppose \(c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = y) \geq c_{1-\alpha}(\|X_{n+1} - AY_n\| |Y_n = z)\),

\[
|h_{+, \delta}(y) - h_{+, \delta}(z)| \leq \int_{c_{1-\alpha}(y) + \delta < \|x\| < c_{1-\alpha}(y) + \delta} |f_{X_{n+1}, A|Y_n}(x|y) - f_{X_{n+1}, A|Y_n}(x|z)|dx
\]

\[
+ \int_{c_{1-\alpha}(y) < \|x\| < c_{1-\alpha}(y)} f_{X_{n+1}, A|Y_n}(x|z)dx + \int_{c_{1-\alpha}(z) + \delta < \|x\| < c_{1-\alpha}(z) + \delta} f_{X_{n+1}, A|Y_n}(x|z)dx \tag{50}
\]

For the first term, notice that conditional density function \(f_{X_{n+1}, A|Y_n}(x|y)\) is continuous, so it is uniformly continuous in compact set \(\|.|^{-1}(c_{1-\alpha}(y), c_{1-\alpha}(y) + \delta) \times \{|z - y|_2 \leq 1\}\), which implies that \(\int_{c_{1-\alpha}(y) + \delta < \|x\| < c_{1-\alpha}(y) + \delta} |f_{X_{n+1}, A|Y_n}(x|y) - f_{X_{n+1}, A|Y_n}(x|z)|dx < \epsilon\) with \(m(.)\) being Lebesgue measure, and

\[
\int_{c_{1-\alpha}(y) < \|x\| < c_{1-\alpha}(y) + \delta} |f_{X_{n+1}, A|Y_n}(x|y) - f_{X_{n+1}, A|Y_n}(x|z)|dx < \epsilon \tag{51}
\]
For the second and third term, from lemma 2 conditional quantile $c_{1-\alpha}(y)$ is continuous, so the integration being small when $y$ and $z$ being sufficiently close. Therefore, $h_+$ is continuous at arbitrary $y \in S_Y$. From assumption in lemma 2 notice that $h_+$ is continuous in compact set $M$, we have $\min_{y \in M} h_+(y) > 0$. Similarly we have $h_-$ being continuous and $\min_{y \in M} h_-(y) > 0$. Notice that conditional quantile $c_{1-\alpha}(y)$ is continuous on $M$, so it has maximum $C$. For any given $1 > \epsilon > 0$, we choose $m$ sufficiently large such that

$$\sup_{y \in M} \|g_m(.|y) - f_{X_{n+1} - AY_n}|Y_n(.|y)\|_{L_p} \leq \frac{\min_{y \in M} \min_{y \in M} h_+(y), \min_{y \in M} h_-(y)}{m^{1/q}}$$

Here $q$ satisfies $1/p + 1/q = 1$ and $r$ is defined. If $\sup_{y \in M} |d_{m, 1-\alpha}(y) - c_{1-\alpha}(|X_{n+1} - AY_n||Y_n = y)| > \epsilon$, then $\exists y_0 \in M$ such that

$$|d_{m, 1-\alpha}(y_0) - c_{1-\alpha}(|X_{n+1} - AY_n||Y_n = y_0)| \geq \epsilon$$

If $d_{m, 1-\alpha}(y_0) \geq c_{1-\alpha}(y_0) + \epsilon$, according to Holder’s inequality and

$$\alpha \leq \int_{\|x\| > c_{1-\alpha}(y_0) + \epsilon} g_m(x|y_0)dx$$

$$\leq \alpha - \int_{c_{1-\alpha}(y_0) \leq \|x\| \leq c_{1-\alpha}(y_0) + \epsilon} f_{X_{n+1} - AY_n}|Y_n(x|y_0) + \int_{\|x\| \leq c_{1-\alpha}(y_0) + \epsilon} (f_{X_{n+1} - AY_n}|Y_n(x|y_0) - g_m(x|y_0))dx$$

$$\leq \alpha - \min_{y \in M} h_+(y) + \|f_{X_{n+1} - AY_n}|Y_n(.|y_0) - g_m(.|y_0)\|_{L_p} m^{1/q}(\|\cdot\|^{-1}(0, 1 - \alpha, + 1))$$

$$\leq \alpha - \min_{y \in M} h_+(y) + \sup_{y \in M} \|f_{|X_{n+1} - AY_n||Y_n = y_0 - } - g_m(.|y_0)\|_{L_p} m^{1/q}(\|\cdot\|^{-1}(0, 1 + C)) < \alpha$$

we get contradiction. On the other hand, if $d_{m, 1-\alpha}(y_0) \leq c_{1-\alpha}(|X_{n+1} - AY_n||Y_n = y_0) - \epsilon$, we have, similarly,

$$\alpha \geq \int_{\|x\| < c_{1-\alpha}(y_0) - \epsilon} g_m(x|y_0)dx$$

$$\geq \alpha + \int_{c_{1-\alpha}(y_0) - \epsilon \leq \|x\| \leq c_{1-\alpha}(y_0) - \epsilon} f_{X_{n+1} - AY_n}|Y_n(x|y_0) - g_m(x|y_0))dx$$

$$\geq \alpha + \min_{y \in M} h_-(y) - \sup_{y \in M} \|f_{X_{n+1} - AY_n}|Y_n(.|y) - g_m(x|y))\|_{L_p} m^{1/q}(\|\cdot\|^{-1}(0, 1 + C)) > \alpha$$

and we get contradiction. Thus, for sufficiently large $m$, [52] is not valid and we prove the result.

**Proof of Theorem** Define $S_Y = \{y|f_Y(y) > 0\}$. Notice that distribution of $Y_n$ is stationary, from dominated convergence theorem, we have

$$\lim_{r \to \infty} Prob(\|Y_n\| > r) = \lim_{r \to \infty} \int_{\|x\| > r} f_Y(x)dx = 0, \lim_{r \to \infty} \int_{\|x\| > r} f_T(x)dx = 0$$

so we can choose sufficiently large $r > 0$ such that $Prob(\|Y_n\| > r) + Prob(f_Y(Y_n) > r) < \epsilon/4$. According to condition 1), for sufficiently large $n$, $Prob(R_n) < \epsilon/4$ and on $R_n$, quantile of $\tilde{f}(x|y)$ is well defined. Set $M = S_Y \cap \|\cdot\|^{-1}(0, r) \cap \{x|f_{Y_n}(x) \leq r\}$ being measurable and has finite Lebesgue measure, from theorem 2.40
in [17], there exists a compact set $K$ such that $K \subset M$ and $m(M) - m(K) < \epsilon/(4r)$, here $m(.)$ denotes Lebesgue measure in $\mathbb{R}^d$. From condition 1, define set $R_n = \left\{ \omega \in \Omega \mid \exists r \in \mathbb{R}^d, y \in K \text{ s.t. } \hat{f}(x|y) < 0 \right\} \cup \left\{ \omega \in \Omega \mid \exists y \in K, f_{R^d}(x|y) \neq 1 \right\}$, we know that $\text{Prob}(R_n) < \epsilon/4$ for sufficiently large $n$ and quantile of $\hat{f}(x|y)$ is well defined for $Y_n \in K$ on $R_n$. For sufficiently large $n$,

$$\text{Prob}(\hat{c}_{1-\alpha}(\|X_{n+1} - \bar{A}Y_n\| | Y_n) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)) > \xi \right)$$

$$\leq \text{Prob}(R_n) + \text{Prob}(\|Y_n\|_2 > r) + \text{Prob}(f_Y(Y_n) > r) + \text{Prob}(Y_n \in M - K)$$

$$+ \text{Prob}(\hat{c}_{1-\alpha}(Y_n) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)) > \xi \cap Y_n \in K))$$

(57)

$$\leq \text{Prob}(R_n) + \frac{\epsilon}{4} + \int_{M-K} f_Y + \text{Prob}(\sup_{y \in K} \{\hat{c}_{1-\alpha}(y) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)\}) > \xi)$$

From condition (55), we choose sufficiently large $n$ such that

$$\text{Prob}(\sup_{y \in K} |\hat{f}(\cdot|y) - f_{X_{n+1} - AY_n} |_{\gamma_n} (\cdot|y)|_{L_p} > \min(\min_{y \in K} h_{+\zeta}(y), \min_{y \in K} h_{-\zeta}(y))) \leq \epsilon/4$$

(58)

definition of $h_{+\zeta}, h_{-\zeta}, C$ see (10) and (12). There exists a measurable set $A_n$ such that $\text{Prob}(A_n) < \epsilon/4$ and for any $\omega \in A_n^c$,

$$\sup_{y \in K} \|\hat{f}(\cdot|y) - f_{X_{n+1} - AY_n}(\cdot|y)|_{L_p} \leq \min(\min_{y \in K} h_{+\zeta}(y), \min_{y \in K} h_{-\zeta}(y)) \leq \epsilon/4$$

(59)

From (52) to (55), we know that $\sup_{y \in K} |\hat{c}_{1-\alpha}(\|X_{n+1} - \bar{A}Y_n\| | Y_n) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)\| | Y_n = y)| \leq \xi$ in $A_n^c$, and $\text{Prob}(\hat{c}_{1-\alpha}(\|X_{n+1} - \bar{A}Y_n\| | Y_n) - c_{1-\alpha}(\|X_{n+1} - AY_n\| | Y_n)) > \xi) \leq \frac{\epsilon}{4} + \int_{M-K} f_Y + \text{Prob}(A_n) \leq \epsilon,$

and we prove the result. \hfill \Box

**Proof of theorem**. According to Riemann-Lebesgue lemma, $f_T, f_S \in L_1$ implies that $\mathcal{F}f_T, \mathcal{F}f_S$ are continuous, and K1) implies that $\|\mathcal{F}K\|_{L_\infty} < \infty$. Fourier inversion formula shows that $\mathcal{F}K_n(t_1, ..., t_d) = \mathcal{F}K(t_1, ..., t_d)$. For given $x, |\mathcal{F}f_T(t)| > 0$ for any $t$ and $\mathcal{F}K_n(t)$ is compactly supported, so

$$|f(t, \omega)| = |\exp(-it^T x) \frac{1}{n} \sum_{k=1}^{n} \exp(it^T Z_k) \frac{\mathcal{F}K_n(t)}{\mathcal{F}f_T(t)}| \leq \frac{\mathcal{F}K_n(t)}{\mathcal{F}f_T(t)} \Rightarrow f(t, \omega) \in L_1(\mathbb{R}^d \times \Omega)$$

(60)

From Fubini theorem, condition (53) and Cauchy inequality, $\mathcal{F}f_S \in L_1 \cap L_2(\mathbb{R}^d)$, so theorem 8.26, 8.29 and lemma 8.34 in [17] shows that inverse Fourier transformation of $\mathcal{F}f_S$ equals $f_S$ almost surely, $f_S \in L_2$ and

$$E \hat{f}_S(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x) \mathcal{F}f_S(t)\mathcal{F}K_n(t)dt = \mathcal{F}^{-1}\mathcal{F}f_S \mathcal{F}K_n(x) = f_S * K_n(x)$$

(61)

First consider $L_2$ norm of $\hat{f}_S$. According to Plancherel theorem and A.6. in [16], since $\mathcal{F}f_S, \mathcal{F}K_n, \mathcal{F}f_S \in$
for some constant $C$ independent of sample size $n$. For the variance of $\tilde{f}_S$, from Fubini’s theorem

$$
\begin{align*}
\text{Var}(\tilde{f}_S) &= \frac{1}{n^2} \sum_{j=1}^{n} \int_{\mathbb{R}^d} \text{E}[|L_h(x - Z_k) - \text{E}f_S(x)|^2] \\
&\leq \frac{1}{n^2} \sum_{j=1}^{n} \int_{\mathbb{R}^d} \text{E}[(L_h(x - Z_k) - \text{E}f_S(x))^2] \\
&\leq (1 + \|FK\|_{L^2})^2 \sum_{j=1}^{d} h_j^{2\rho_j} \exp(-2\alpha_j h_j^{-\rho_j})
\end{align*}
$$

(63)

For the first term in (63), according to Plancherel theorem and lemma 1 in [16], we have $\|L_h(-Z_k) - \text{E}f_S\|_{L^2} = \frac{1}{(2\pi)^d} \|\mathcal{F}f_S \mathcal{F}K_h \|_{L^2}$ and correspondingly

$$
\begin{align*}
\int_{\mathbb{R}^d} \text{E}[|L_h(x - Z_k) - \text{E}f_S(x)|^2] &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \text{E}[|\mathcal{F}K_h(x)|^2] \exp(x \cdot \mathcal{F}f_S - \mathcal{F}f_S) \text{d}x \\
&\leq \frac{\|FK\|_{L^2}}{c^2} \prod_{j=1}^{d} \int_{-\infty}^{\infty} x_j^{2\rho_j} \exp(2\alpha_j |x_j|^{\rho_j}) \text{d}x_j \\
&\leq C \prod_{j=1}^{d} h_j^{-2\rho_j - 1 + 2\rho_j - \rho_j} \exp(2\alpha_j h_j^{\rho_j})
\end{align*}
$$

(64)

with constant $C$ independent of $h_j$. For second term in (63), with given $x \in \mathbb{R}^d$ and arbitrary $k \in \mathbb{Z}^+$, we let $\mathcal{G}_0, \mathcal{G}_k$ respectively being $\sigma$-algebra generated by $\sin(x^T Z_j), \cos(x^T Z_j), j \leq 0$ and $\sin(x^T Z_j), \cos(x^T Z_j), j \geq k$ and $\mathcal{H}_0, \mathcal{H}_k$ being $\sigma$-algebra generated by $Z_j, j \leq 0$ and $Z_j, j \geq k$. Notice that function $\sin(x^T \cdot), \cos(x^T \cdot)$ are continuous and thus measurable in $\mathbb{R}^d$, so for any Borel set $B \subset \mathbb{C}$ and any $j \leq 0$, $\sin(x^T \cdot)^{-1}(B)$ is Borel measurable in $\mathbb{R}^d$ and $\{\omega \in \Omega | \sin(x^T Z_j) \in B\} = \{\omega \in \Omega | \sin(x^T \cdot)^{-1}(B)\} \in \mathcal{H}_0$, so $\sin(x^T Z_j)$ is measurable in $\mathcal{H}_0$ for any $j \leq 0$. Similarly $\cos(x^T Z_j)$ is measurable in $\mathcal{H}_0$ for any $j \leq 0$, which implies that $\mathcal{G}_0 \subset \mathcal{H}_0$. Use the same technique we can show that $\mathcal{G}_k \subset \mathcal{H}_k$. Therefore, define $\alpha_\mathcal{G}(k)$ being $\alpha$ mixing coefficient generated by $\sigma$-algebras $\mathcal{G}_k, k \in \mathbb{Z}^+$, we have

$$
\alpha_\mathcal{G}(k) = \sup \{|\text{Prob}(A \cap B) - \text{Prob}(A) \text{Prob}(B)| | A \in \mathcal{G}_0, B \in \mathcal{G}_k\} \leq \alpha_x(k)
$$

(65)

and correspondingly from lemma 3 in [28], we have

$$
\begin{align*}
|\text{E} \exp(\imath x^T Z_k - \imath y^T Z_j) - \mathcal{F}f(x) \mathcal{F}f(x)| &\leq |\text{E} \cos(x^T Z_k) \cos(x^T Z_j) - \text{E} \cos(x^T Z_k) \text{E} \cos(x^T Z_j)| \\
+ |\text{E} \sin(x^T Z_k) \sin(x^T Z_j) - \text{E} \sin(x^T Z_k) \text{E} \sin(x^T Z_j)| &\leq 16\alpha_\mathcal{G}(|k - j|) \leq 16\alpha_x(k - j)
\end{align*}
$$

(66)
On the other hand, for $k \neq j$, from Plancherel’s theorem and (63),

\[
L_h(x - Z_k)L_h(x - Z_j) + L_h(x - Z_j)L_h(x - Z_k) = \frac{1}{2}((L_h(x - Z_k) + L_h(x - Z_j))^2 - (L_h(x - Z_k) - L_h(x - Z_j))^2)
\]

\[
\Rightarrow \int_{\mathbb{R}^d} L_h(x - Z_k)L_h(x - Z_j) + L_h(x - Z_j)L_h(x - Z_k) = \frac{1}{2}\left(\|F^{-1}\left(\exp(it^T Z_k)\frac{FK_h}{\mathcal{F}T} + \exp(it^T Z_j)\frac{FK_h}{\mathcal{F}T}\right)\|_2^2 - \|F^{-1}\left(\exp(it^T Z_k)\frac{FK_h}{\mathcal{F}T} - \exp(it^T Z_j)\frac{FK_h}{\mathcal{F}T}\right)\|_2^2\right)
\]

\[
= \frac{1}{2} \left(\|\exp(it^T Z_k)\frac{FK_h}{\mathcal{F}T}\|_2^2 - \|\exp(it^T Z_j)\frac{FK_h}{\mathcal{F}T}\|_2^2\right)
\]

\[
= \frac{1}{2} \left(\|\exp(it^T Z_k)\frac{FK_h}{\mathcal{F}T}\|_2^2 - \|\exp(it^T Z_j)\frac{FK_h}{\mathcal{F}T}\|_2^2\right)
\]

\[
= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} |\frac{FK_h(t)}{\mathcal{F}T(t)}|^2 (\exp(it^T Z_k - it^T Z_j) + \exp(it^T Z_j - it^T Z_k) - 2Ff_z(t)\mathcal{F}f_z(t))dt + 2\|f_s\*K_h\|_2^2
\]

(67)

Combine (61), (66) and (67), second term of (63) satisfies

\[
\frac{1}{n^2} E \sum_{j,k=1,k \neq j}^{n} \int_{\mathbb{R}^d} (L_h(x - Z_k) - \mathbb{E}f_s(x))(L_h(x - Z_j) - \mathbb{E}f_s(x))
\]

\[
= \frac{1}{2n^2} E \sum_{j=1,k \neq j}^{n} \int_{\mathbb{R}^d} L_h(x - Z_k)L_h(x - Z_j) + L_h(x - Z_j)L_h(x - Z_k) - 2|\mathbb{E}f_s(x)|^2 dx
\]

\[
= \frac{1}{2n^2} \sum_{j,k=1,k \neq j}^{n} \int_{\mathbb{R}^d} |\frac{FK_h(x)}{\mathcal{F}T(x)}|^2 (\exp(ix^T Z_k - ix^T Z_j) + \exp(ix^T Z_j - ix^T Z_k) - 2Ff_z(x)\mathcal{F}f_z(x)) dx
\]

\[
\leq \frac{1}{n^2} \sum_{j,k=1,k \neq j}^{n} \int_{\mathbb{R}^d} |\frac{FK_h(x)}{\mathcal{F}T(x)}|^2 (\exp(ix^T Z_k - ix^T Z_j) + \exp(ix^T Z_j - ix^T Z_k) - 2Ff_z(x)\mathcal{F}f_z(x)) dx
\]

\[
\leq \frac{1}{n} \sum_{j,k=1,k \neq j}^{n} \alpha_{z}(k-j) \int_{\mathbb{R}^d} |\frac{FK_h(x)}{\mathcal{F}T(x)}|^2
\]

\[
= \frac{1}{n} \sum_{j,k=1,k \neq j}^{n} \alpha_{z}(k-j) \sum_{s=1}^{n-1} \left(1 - \frac{|s|}{n}\right) \alpha_{z}(s)
\]

\[
\leq C \prod_{j=1}^{d} h_j^{-2\beta_j-1+\rho_j} \exp(2\alpha_j a^\rho_j h_j^{-\rho_j})
\]

(68)

for some constant $C$ independent of $n$ and $h = (h_1, ..., h_d)$. Combine (62), (64) and (65), there exists constant $C$ such that

\[
\mathbb{E}\|\tilde{f}_s - f_s\|_2^2 \leq 2\mathbb{E}\|\tilde{f}_s - \mathbb{E}\tilde{f}_s\|_2^2 + 2\mathbb{E}\|\mathbb{E}\tilde{f}_s - f_s\|_2^2
\]

\[
\leq C \left(\frac{1}{n} \prod_{j=1}^{d} h_j^{-2\beta_j-1+\rho_j} \exp(2\alpha_j a^\rho_j h_j^{-\rho_j}) + \sum_{j=1}^{d} h_j^{2\beta_j} \exp(-2\alpha_j h_j^{-\rho_j})\right)
\]

(69)

Define $\tilde{g}_s = \max(\text{Re}\tilde{f}_s, 0)$ since $f_s \geq 0$, we have

\[
|\tilde{f}_s - f_s|^2 = |\text{Re}\tilde{f}_s - f_s|^2 + |\text{Im}\tilde{f}_s|^2 \geq |\tilde{g}_s - f_s|^2
\]

\[
\Rightarrow \mathbb{E}\|\tilde{g}_s - f_s\|_2^2 \leq C \left(\frac{1}{n} \prod_{j=1}^{d} h_j^{-2\beta_j-1+\rho_j} \exp(2\alpha_j a^\rho_j h_j^{-\rho_j}) + \sum_{j=1}^{d} h_j^{2\beta_j} \exp(-2\alpha_j h_j^{-\rho_j})\right)
\]

(70)

with constant $C$ defined in (69). From Cauchy inequality, we have, for chosen bandwidth $s = s(n) = (s_1, ..., s_d)$
tending to infinity and some constant $C$,

$$E|\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \tilde{g}_S - 1| ^2 \leq 2E(\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} |\tilde{g}_S - f_S| dx)^2 + 2\int_{([-s_1, s_1] \times \ldots \times [-s_d, s_d])^c} f_S(x) dx)^2$$

$$\leq 2^{d+1} s_1 \ldots s_d \times E\|\tilde{g}_S - f_S\|_2^2 + 2\int_{([s_1, s_1] \times \ldots \times [s_d, s_d])^c} f_S(x) dx)^2$$

$$\leq C \left( \frac{1}{n} \prod_{j=1}^{d} s_j h_j^{-2b_j - 1 + p_j} \exp(2a_j \alpha_{\rho_j} h_j^{-\rho_j}) + \prod_{j=1}^{d} h_j^{2b_j} s_1 \ldots s_d \exp(-2a_j h_j^{-\rho_j}) \right) + 2\eta^2$$

(71)

In particular, from Markov inequality, for given $1 > \eta, \epsilon > 0$ and sufficiently large $n$, with bandwidth $s = s(n) = (s_1, \ldots, s_d)$ and $s_i(n) \to \infty$ for $i = 1, 2, \ldots, d$, we have $\int_{([-s_1, s_1] \times \ldots \times [-s_d, s_d])^c} f_S(x) dx < \eta$.

$$\text{Prob}(\|\text{max}(\text{Re} S, 0)\|_1 \geq \epsilon) \leq \frac{1}{\epsilon^2} E(\int_{\mathbb{R}^d} \text{max}(\text{Re} S, 0) \prod_{j=1}^{d} [-s_j, s_j] - f_S| dx)^2$$

$$\leq \frac{2}{\epsilon^2} \left( E(\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \text{max}(\text{Re} S, 0) \prod_{j=1}^{d} [-s_j, s_j] - f_S dx)^2 \right)$$

$$\leq C \left( \frac{1}{n} \prod_{j=1}^{d} s_j h_j^{-2b_j - 1 + p_j} \exp(2a_j \alpha_{\rho_j} h_j^{-\rho_j}) + \prod_{j=1}^{d} h_j^{2b_j} s_1 \ldots s_d \exp(-2a_j h_j^{-\rho_j}) \right) + 2\epsilon^2$$

(72)

which implies that $\|\text{max}(\text{Re} S, 0)\|_1 \prod_{j=1}^{d} [-s_j, s_j] - f_S| \to 0$ as long as bandwidth $s = s(n) = (s_1, \ldots, s_d)$ with $s_i \to \infty$ for $i = 1, 2, \ldots, d$ and (56) holds. Finally,

$$\|\tilde{S} - f_S\|_1 \leq \frac{\|\text{max}(\text{Re} S, 0)\|_1 \prod_{j=1}^{d} [-s_j, s_j] - f_S\|_1}{\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \text{max}(\text{Re} S, 0) dx} + \frac{\|\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \text{max}(\text{Re} S, 0) \prod_{j=1}^{d} [-s_j, s_j] - f_S dx\|_1}{\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \text{max}(\text{Re} S, 0) dx}$$

(74)

from (72) and (73), we prove the result.

Proof of lemma. We use $\|\cdot\|_F$ to denote Frobenius norm $\|A\|_F = \sqrt{\text{tr}(AA^T)}$, first notice that for any given $k$,

$$Y_k Y_k^T = A^2 X_{k-2} X_{k-2}^T + \epsilon_k X_{k-2}^T + \eta_k X_{k-2}^T + A^2 X_{k-2} \eta_k \eta_k - \epsilon_k \eta_k - \eta_k \eta_k$$

$$Y_k Y_k^T = A^2 X_{k-1} X_{k-1}^T + \epsilon_k X_{k-1}^T + \eta_k X_{k-1}^T + A^2 X_{k-1} \eta_k \eta_k - \epsilon_k \eta_k - \eta_k \eta_k$$

(75)
From condition $A3)$, notice that $\|X_k\|_2 \leq \sum_{j=0}^{\infty} \|A\|_2^j \|\epsilon_{k-j}\|_2$, so

$$E\|X_k\|_2^2 \leq E \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} \|A\|_2^{j+r+s} E(||\epsilon_{k-j}\|_2 \|\epsilon_{k-j-1}\|_2 \|\epsilon_{k-j-2}\|_2 \|\epsilon_{k-j-3}\|_2)$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \sum_{r=0}^{\infty} \|A\|_2^{j+r+s} E(||\epsilon_{k-j}\|_2^2 = E\|\epsilon_{k-j}\|_2^2 \left(\sum_{j=0}^{\infty} \|A\|_2^j \right)^4 < \infty$$

(76)

and from Fubini's theorem

$$S = EX_kX_k^T = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} E A^j \epsilon_{k-j}^T \epsilon_{k-j} - A^T = \sum_{j=0}^{\infty} A^j \Sigma A^j$$

(77)

exists. According to [75].

$$\| \frac{1}{n-2} \sum_{k=1}^{n} Y_k Y_{k-2} - A^2 S \|_2 \leq \| A \|_2 \left( \frac{1}{n-2} \sum_{k=1}^{n} X_k X_k^T - S \right) + \| A \|_2 \left( \frac{1}{n-2} \sum_{k=1}^{n} \epsilon_k + X_k \right) \|_2 + \| A \|_2 \left( \frac{1}{n-2} \sum_{k=1}^{n} \epsilon_k + X_k \right) \|_2$$

(78)

Notice that

$$E \| X_k \|_2^4 - \| A \|_2^2 \left( \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} E(A^j \epsilon_{k-j}^T \epsilon_{k-j} - A^T) \right)^2 \leq \| A \|_2^2 \left( \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} E(A^j \epsilon_{k-j}^T \epsilon_{k-j} - A^T) \right)^2 - \| A \|_2^2 \leq \frac{1}{n-2} \| A \|_2^2 \left( \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} E(A^j \epsilon_{k-j}^T \epsilon_{k-j} - A^T) \right)^2$$

(79)

According to Cauchy inequality and theorem 5.6.2 in [29],

$$E \left( \sum_{i=0}^{k-j-1} X_i^T A^j \epsilon_{k-j} + X_i^T A^{k-j} X_i \right)^2 - \| A \|_2^2 \left( \sum_{j=0}^{\infty} \| A^j \|_2^2 \right)^2$$

$$\leq \| A \|_2^2 \left( \sum_{j=0}^{\infty} \| A^j \|_2^2 \right)^2$$

(80)
for any $n \geq 3$. Similarly we have
\begin{align*}
\mathbb{E}\| \frac{1}{n} \sum_{k=1}^{n-2} \epsilon_{k+1}^T X_k \|_2^2 &\leq \frac{1}{(n-2)^2} \sum_{k=1}^{n-2} \sum_{j=1}^{n-2} \mathbb{E} X_k^T X_j \epsilon_{j+1} \epsilon_{k+1} = \frac{1}{n-2} \mathbb{E}\| X_k \|_2^2 \mathbb{E}\| \epsilon_k \|_2^2 \\
\mathbb{E}\| \frac{1}{n-2} \sum_{k=1}^{n-2} X_k \eta_{k+2} \|_2^2 &\leq \frac{1}{(n-2)^2} \sum_{k=1}^{n-2} \sum_{j=1}^{n-2} \mathbb{E} X_k^T X_j \eta_{j+2} \eta_{k+2} = \frac{1}{n-2} \mathbb{E}\| \epsilon_k \|_2^2 \mathbb{E}\| \eta_k \|_2^2 \\
\mathbb{E}\| \frac{1}{n-2} \sum_{k=1}^{n-2} \eta_{k+2} \eta_k \|_2^2 &\leq \frac{1}{(n-2)^2} \sum_{k=1}^{n-2} \sum_{j=1}^{n-2} \mathbb{E} \eta_{j+2} \eta_k \eta_{j+2} \eta_{k+2} = \frac{1}{n-2} \mathbb{E}\| \eta_k \|_2^2 \mathbb{E}\| \eta_k \|_2^2
\end{align*}

Thus there exists a constant $C$ such that for any $s > 0$ and any $n > 3$,
\begin{align*}
\text{Prob} \left( \| \frac{1}{n} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A^2 S \|_2 \geq s \right) &\leq \frac{\mathbb{E}\| \frac{1}{n} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A^2 S \|_2^2}{s^2} \leq \frac{C}{(n-2)s^2}
\end{align*}

Apply corollary 6.3.4 in [29] to matrix $A \Sigma^2 A^T$, which is normal and positive definite, let $s$ in \((82)\) being sufficiently small and $\| \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma \|_2 < s$, then for any eigenvalue $\hat{\lambda}$ of $\left( \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T \right)^T$, there exists an eigenvalue $\lambda$ of $A \Sigma^2 A^T$ such that
\begin{align*}
\| \hat{\lambda} - \lambda \| \leq A \Sigma^2 A^T - \left( \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T \right)^T \leq \| \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma \|_2 + \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma \|_2 < \frac{\min_{\lambda \in \sigma(A \Sigma^2 A^T)} \lambda}{4}
\end{align*}

which implies that $\frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T$ being invertible. From (5.8.3) in [29], for sufficiently small $s$ such that $\| \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma \|_2 < s \Rightarrow \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T$ being invertible and $1 - \| (A \Sigma)^{-1} \|_2 > 1/2$,
\begin{align*}
\text{Prob}(\| \hat{A} - A \| \geq s) &\leq \text{Prob}(| \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma |_2 \geq s) \\
&+ \text{Prob}(\frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A^2 \Sigma |_2 \left( \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T \right)^{-1} \|_2 \geq \frac{s}{2}) \\
&+ \text{Prob}(\| A^2 \Sigma \|_2 \left( \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T \right)^{-1} - (A \Sigma)^{-1} \|_2 \|_2 \geq \frac{s}{2}) \\
&\leq \text{Prob}(\| \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma \|_2 \geq s) \\
&+ \text{Prob}(\| \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A^2 \Sigma \|_2 \left( \| (A \Sigma)^{-1} \|_2 + \frac{\| (A \Sigma)^{-1} \|_2^2 S}{1 - \| (A \Sigma)^{-1} \|_2^2} \right) \geq \frac{s}{2}) \\
&+ \text{Prob}(\| A^2 \Sigma \|_2 \left( \| (A \Sigma)^{-1} \|_2 \|_2 s \right) \leq \text{Prob}(| \frac{1}{n-2} \sum_{k=3}^{n} Y_k Y_{k-1}^T - A \Sigma |_2 \geq \frac{s}{2}) \leq \frac{C}{(n-2)s^2}
\end{align*}

for some constant $C$ and arbitrary $n \geq 3$, and we prove the result.
Proof of theorem [3] will be divided into several parts. We first give a lemma.

**Lemma 6.** Suppose data \( Y_1, ..., Y_n \) are gathered from stationary model [6] and conditions A1) to A3) are satisfied. If estimated density \( \hat{f} \) and \( f \) satisfy

1) \( \text{Prob}(\exists x \in \mathbb{R}^d, \hat{f}(x) < 0) + \text{Prob}(\exists x \in \mathbb{R}^d, \hat{f}(x) < 0) \to 0 \) and \( \text{Prob}(\hat{f}_{\mathbb{R}^d}(x)dx \neq 1) + \text{Prob}(\hat{f}_{\mathbb{R}^d}(x)dx \neq 1) \to 0 \)

2) \( \| \hat{f} - f \|_{L_1} \xrightarrow{n \to \infty} 0, \| \hat{f} - f \|_{L_1} \xrightarrow{n \to \infty} 0, \| \hat{f} \star f_n - f \|_{L_\infty} \xrightarrow{n \to \infty} 0 \) \( \| \hat{f} \star f_n - f \|_{L_\infty} \xrightarrow{n \to \infty} 0 \)

then the estimated density \( f_{X_{n+1}^{*} - \hat{A}Y_n}|Y_1, ..., Y_n = \epsilon_{X_{n+1}^{*} - \hat{A}Y_n}|Y_1, ..., Y_n \) generated in bootstrap algorithm [3] satisfies condition 1) and 2) in theorem [3].

**Proof of lemma [6].** From condition A2), Cauchy’s inequality and [113], we have \( f \in L_1 \Rightarrow f \in L_1 \), combine with theorem 8.26 in [17], there exists a version of density \( f \) which is continuous and \( \| f \|_{L_\infty} \leq \frac{1}{12 \pi} \| \mathcal{F}f \|_{L_1} < \infty \). According to theorem 8.8 in [17], notice \( f_n \in L_1 \) and \( f \in L_\infty \), so \( f = f \star f_n \) exists uniformly continuous. Define set \( S_Y = \{ y \in \mathbb{R}^d | f_y(y) > 0 \} \), from [29], we have for any \( (x', y') \in \mathbb{R}^d \times S_Y \) such that \( \| x - x' \|_2, \| y - y' \|_2 \) being sufficiently small and \( (x', y') \in \mathbb{R}^d \times S_Y \),

\[
|f_{X_n^{*} - AY_n}(x, y) - f_{X_n^{*} - AY_n}(x', y')| \\
\leq \frac{1}{|\det(A)|} \int_{\mathbb{R}^d} |f_y(A^{-1}(t - x)) - f_y(A^{-1}(t - x'))|f_X(y + A^{-1}(t - x))f(t)dt \\
+ \frac{1}{|\det(A)|} \int_{\mathbb{R}^d} |f_X(y + A^{-1}(t - x)) - f_X(y' + A^{-1}(t - x'))|f_y(A^{-1}(t - x'))f(t)dt \\
\leq \frac{1}{|\det(A)|} \| f_X \|_{L_\infty} \| f_y \|_{L_\infty} \| f_y(A^{-1}.) - f_y(A^{-1}(x' - x + y' - y + A^{-1}.)\|_{L_1} \\
+ \frac{1}{|\det(A)|} \| f_y \|_{L_\infty} \| f_y \|_{L_\infty} \| f_X(A^{-1}.) - f_X(A^{-1}(x' - x + y' - y + A^{-1}.)\|_{L_1}
\]

From Proposition 8.5 in [17] we know that \( f_{X_n^{*} - AY_n}(x, y) \) is continuous in \( \mathbb{R}^d \times S_Y \). From lemma [4] since \( f_Y(y) \) is continuous on \( y \in S_Y \) and \( f_{X_n^{*} - AY_n}(x, y) \) is continuous, so \( f_{X_n^{*} - AY_n}(x|y) \) is continuous on \( \mathbb{R}^d \times S_Y \).

According to lemma [5] and continuity of determinant, for sufficiently large \( n \) with high probability \( \hat{A} \) is non-singular and \( |\det(\hat{A})| > 0. \| \hat{f} \star f_n - f \|_{L_\infty} \xrightarrow{n \to \infty} 0 \) and compact set \( K \subset S_Y \), so \( \min_{y \in K} f_Y > 0 \) and for sufficiently large \( n \) with high probability, \( \| \hat{f} \star f_n - f \|_{L_\infty} < \min_{y \in K} f_Y/4 \), which implies that

\[
0 < \frac{3}{4} \min_{y \in K} f_Y \leq \hat{f} \star f_n \leq \| f \|_{L_\infty} + \min_{y \in K} f_Y/4 < \infty
\]

for almost surely \( y \in K \). From Proposition 8.8 in [17] and \( \| f_n \|_{L_\infty} < \infty \), \( \hat{f} \in L_1 \) we know that [87] holds for any \( y \in K \) and sufficiently large \( n \). From change of variable theorem, when \( |\det(A)| + 1 > |\det(\hat{A})| > |\det(A)/2 > 0 \) and condition [87] holds, at \( \omega \in \Omega \) where \( \hat{f} \) and \( \hat{f} \) are non-negative and have integration 1, density of \( X_{n+1}^{*} - \hat{A}Y_n|Y_n \) is given by

\[
f_{X_{n+1}^{*} - \hat{A}Y_n|Y_n}(x|y) = \frac{1}{|\det(A)| \times \hat{f}(y)} \int_{\mathbb{R}^d} \hat{f}(t) \hat{f}(y - \hat{A}^{-1}(t - x))f_y(\hat{A}^{-1}(t - x))dt
\]

(88)
For any $y \in K$, because $\hat{f}_X, \hat{f}_r \geq 0$ and

$$\int_{\mathbb{R}^d} f_{X_n+1-A_Y|Y_n}(x|y)dx = \int_{\mathbb{R}^d} \hat{f}_c(t)dt = 1$$

(89)

According to condition 1), condition 1) in theorem holds.

On the other hand, for any $y \in K$,

$$\|f_{X_n+1-A_Y|Y_n}(\cdot|y) - f_{X_n+1-A_Y|Y_n}(\cdot|y)\|_{L_1} = \|\frac{\hat{f}_c \ast (\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1})))}{|\det(\hat{A})| \times \hat{f}_c \ast f_\eta(y)} - \frac{f_c \ast (f_X(y + A^{-1})f_\eta(-A^{-1})))}{|\det(A)| \times f_c \ast f_\eta(y)}\|_{L_1}$$

$$\leq \frac{1}{|\det(\hat{A})| \times \hat{f}_c \ast f_\eta(y)} - \frac{1}{|\det(A)| \times f_c \ast f_\eta(y)} \|\hat{f}_c \ast (\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1}))\|_{L_1}$$

$$\leq \frac{1}{|\det(A)| \times f_c \ast f_\eta(y)} \|\hat{f}_c - f_c\|_{L_1} \ast (\hat{f}_X(y + A^{-1})f_\eta(-A^{-1}))\|_{L_1}$$

$$\|f_c \ast (\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1})) - f_X(y + A^{-1})f_\eta(-A^{-1})\|_{L_1}$$

$$\leq \|f_c\|_{L_1} \|\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1})) - f_X(y + A^{-1})f_\eta(-A^{-1})\|_{L_1}$$

$$\leq \|f_c\|_{L_1} \|\hat{f}_X - f_X\|_{L_1} \|\det(\hat{A})\| \frac{\text{Prob}_{n \to \infty}}{0}$$

(90)

From Young's inequality, at $\omega \in \Omega$ such that condition 1) holds, for sufficiently large $n$,

$$\|\hat{f}_c - f_c\|_{L_1} \leq \|\hat{f}_c\|_{L_1} \ast \||\hat{f}_c\|_{L_1} \ast \int_{\mathbb{R}^d} f_X(x)dx$$

(91)

$$\|f_c \ast (\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1}))) - f_X(y + A^{-1})f_\eta(-A^{-1})\|_{L_1}$$

$$\leq \|f_c\|_{L_1} \|\hat{f}_X(y + \hat{A}^{-1})f_\eta(-\hat{A}^{-1}))) - f_X(y + A^{-1})f_\eta(-A^{-1})\|_{L_1}$$

$$\leq \|f_c\|_{L_1} \|\hat{f}_X - f_X\|_{L_1} \|\det(\hat{A})\| \frac{\text{Prob}_{n \to \infty}}{0}$$

(92)

Notice that $f_\eta, f_X \in L_1$, from Proposition 7.9 in [17], for any $\epsilon > 0$, we can find function $h_\eta, h_X$ being continuous with compact support, and thus uniform continuous such that $\|f_\eta - h_\eta\|_{L_1} < \epsilon/4, \|f_X - h_X\|_{L_1} < \epsilon/4$. For $h_\eta$ has compact support, there exists a constant $r$ such that $\text{supp} h_\eta \subset \{x|x_2 \leq r\}$. For $A - \hat{A}\|_2$ being sufficiently small such that corollary 6.3.4 in [20] implies that minimum eigenvalue of matrix $\hat{A}^T \hat{A}$ and $A^T A$, $\hat{\lambda}_{\min}, \lambda_{\min}$ satisfies $\hat{\lambda}_{\min} \geq \lambda_{\min}/2$. Let $\lambda_{\max}$ be largest eigenvalue of $A^T A \hat{z} = A^{-1} \hat{A}x$, we have

$$\sqrt{\lambda_{\max}} \|z\|_2 \geq \|A\|_2 = \|\hat{A}\|_2 \geq \sqrt{\lambda_{\min}} \|x\|_2 \Rightarrow \|A^{-1} \hat{A}x\|_2 \geq \sqrt{\lambda_{\min}/2\lambda_{\max}} \|x\|_2$$

(93)

(94)
correspondingly \( \supp h_\eta(A^{-1} \tilde{A}) \subset \{ x \| A^{-1} \tilde{A} x \| \leq r \} \subset \{ x \| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} r} \} \), which implies that \( \supp h_\eta \cup \supp h_\eta(A^{-1} \tilde{A}) \subset \{ x \| x \| \leq \max(1, \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} r}) \} \).

\[
\int_{\mathbb{R}^d} |f_\eta(x) - f_\eta(A^{-1} \tilde{A} x)| dx \lesssim \| f_\eta - h_\eta \|_{L_1} + \int_{\mathbb{R}^d} |h_\eta(x) - h_\eta(A^{-1} \tilde{A} x)| dx + \left( \frac{\det(A)}{\det(A)} \right) \| f_\eta - h_\eta \|_{L_1} \leq \left( 1 + \left( \frac{\det(A)}{\det(A)} \right) \right) \frac{\epsilon}{4} + \int_{\| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} r}} |h_\eta(x) - h_\eta(A^{-1} \tilde{A} x)| dx
\]

(95)

Notice that \( \| x - A^{-1} \tilde{A} x \| \leq \| A^{-1} \|_2 \| A - \tilde{A} \|_2 \| x \| \), with \( \| A - \tilde{A} \|_2 \) sufficiently small \( \epsilon \) is less than \( \left( 1 + \left( \frac{\det(A)}{\det(A)} \right) \right) \frac{\epsilon}{4} + \epsilon \times m \left( \| x \| \leq \max(1, \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} r}) \right) \) from continuity of determinant, we have \( |\det(\tilde{A})| > |\det(A)/2 \) for \( \| A - \tilde{A} \|_2 \) being sufficiently small, so

\[
\int_{\mathbb{R}^d} |f_\eta(x) - f_\eta(A^{-1} \tilde{A} x)| dx \xrightarrow{\text{Prob}} 0 \quad (96)
\]

Similarly because \( h_Y \) is compactly support and uniformly continuous, there exists constant \( s \) such that \( \supp h_Y \subset \{ \| x \| \leq s \} \), and

\[
\supp h_Y(y - A^{-1} \tilde{A} y + A^{-1} \tilde{A}) \subset \{ x \| A^{-1} \tilde{A} (A^{-1} Ay - y + x) \| \leq s \}
\]

\[
\subset \{ x \| A^{-1} Ay - y + x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s} \}
\]

(97)

For \( y \in K \) and \( K \) is a compact set, so there exists a constant \( t > 0 \) such that \( K \subset \{ x \| x \| \leq t \} \), and from (5.8.3) in [29],

\[
\left\{ x \| A^{-1} Ay - y + x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s} \right\} \subset \left\{ x \| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s + \| y \|_2 + \| A^{-1} \|_2 \| A \|_2 \| y \|_2} \right\}
\]

\[
\subset \left\{ x \| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s + t + \left( \| A^{-1} \|_2 + \| A^{-1} \|_2 \| \tilde{A} - A \|_2 \| A \|_2 t \right) \| A \|_2 t} \right\}
\]

(98)

For sufficiently large \( n, \| A^{-1} \|_2 \| A - \tilde{A} \|_2 < 1/2 \) with high probability and correspondingly

\[
\int_{\mathbb{R}^d} |f_Y(x) - f_Y(A^{-1} \tilde{A} x + y - A^{-1} \tilde{A} y)| dx \leq \left( 1 + \left( \frac{\det(A)}{\det(A)} \right) \right) \| f_Y - h_Y \|_{L_1} + \int_{\| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s + t + (\| A^{-1} \|_2 + 1/2) \| A \|_2 t} |h_Y(x) - h_Y(A^{-1} \tilde{A} x + y - A^{-1} \tilde{A} y)| dx
\]

(99)

Notice that \( \| x - A^{-1} \tilde{A} x + y - A^{-1} \tilde{A} y \|_2 \leq \| A^{-1} \|_2 \| A - \tilde{A} \|_2 (\| x \|_2 + \| y \|_2) \leq \| A^{-1} \|_2 \| A - \tilde{A} \|_2 (\sqrt{2\lambda_{\max}} s + t + (\| A^{-1} \|_2 + 1/2) \| A \|_2 t + t) \) and \( h_Y \) be uniform continuous, for \( \| A - \tilde{A} \|_2 \) being sufficiently small,

\[
\int_{\| x \| \leq \sqrt{\frac{2\lambda_{\max}}{\lambda_{\min}} s + t + (\| A^{-1} \|_2 + 1/2) \| A \|_2 t} |h_Y(x) - h_Y(A^{-1} \tilde{A} x + y - A^{-1} \tilde{A} y)| dx
\]

\[
< \epsilon \times m(\| x \|_2) \leq \frac{2\lambda_{\max}}{\lambda_{\min}} s + t + (\| A^{-1} \|_2 + 1/2) \| A \|_2 t) \Rightarrow \int_{\mathbb{R}^d} |f_Y(x) - f_Y(A^{-1} \tilde{A} x + y - A^{-1} \tilde{A} y)| dx \xrightarrow{\text{Prob}} 0 \quad (100)
\]
uniformly for any \( y \in K \). Finally, notice that for sufficiently large \( n \) with high probability

\[
\frac{1}{|\det(\hat{A})| \times f_X * f_\eta(y)} - \frac{1}{|\det(\hat{A})| \times f_X * f_\eta(y)} \leq \frac{|\det(\hat{A}) - \det(A)||f_X * f_\eta(y) + |\det(\hat{A})||\hat{f}_X \ast f_\eta(y) - f_X * f_\eta(y)|}{|\det(\hat{A})| \times f_X * f_\eta(y) \times |\det(\hat{A})| \times f_X * f_\eta(y)} \leq 8\frac{|\det(\hat{A}) - \det(A)||f_X\|_\infty + |\det(\hat{A})||\hat{f}_X \ast f_\eta - f_Y\|_\infty}{3 \times |\det(\hat{A})|^2 \times (\min_{\eta \in K} f_Y)^2} \xrightarrow{n \to \infty} 0
\]

(101)

uniformly in \( K \) and

\[
\|\hat{f}_X \ast (\hat{f}_X(y + \hat{A}^{-1}) f_\eta(-\hat{A}^{-1}))\|_{L_1} \leq \|(\hat{f}_X - f_\eta) \ast (\hat{f}_X(y + \hat{A}^{-1}) f_\eta(-\hat{A}^{-1}))\|_{L_1} + \|f_\eta \ast (\hat{f}_X(y + \hat{A}^{-1}) f_\eta(-\hat{A}^{-1}))\|_{L_1} \leq \|f_\eta\|_{L_\infty} \|\det(\hat{A})\|
\]

(102)

From (91) and (92), first two terms in (102) tends to 0 uniformly in \( K \) in probability, so they are bounded with high probability uniformly in \( K \), for the third term from Young’s inequality,

\[
\|f_\eta \ast f_X(y + \hat{A}^{-1}) f_\eta(-\hat{A}^{-1})\|_{L_1} \leq \|f_\eta\|_{L_1} \|f_X(y + \hat{A}^{-1}) f_\eta(-\hat{A}^{-1})\|_{L_1} \leq \|f_\eta\|_{L_\infty} \|\det(\hat{A})\|
\]

(103)

is uniformly bounded in \( K \) with high probability. Combine with (90) to (92), (96), (100) to (103), we prove that

\[
\sup_{y \in K} \|\hat{f}_X \ast f_\eta(y)\|_{L_1} \xrightarrow{\text{Prob}} 0
\]

(104)

\[
\text{Combinewithlemma6andtheorem1,wehaveconditionalquantilefunction} \ c_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_1, \ldots, Y_n) \text{converges to} \ c_{1-\alpha}(\|X_{n+1} - A Y_n\| Y_n) \text{in probability. Second lemma shows that, if conditions in lemma6 holds, then the underlying conditional quantile function of bootstrap random variable} \ |X_{n+1} - \hat{A} Y_n| Y_1, \ldots, Y_n \text{converges to the real quantile function} \ c_{1-\alpha}(\|X_{n+1} - A Y_n\| Y_n) \text{in probability.}
\]

**Lemma 7.** Suppose conditions in lemma6 and \( \alpha \in (0, 1) \) is chosen so that \( A4 \) is satisfied, then we have

\[
|c_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_n) - \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_1, \ldots, Y_n)| \xrightarrow{\text{Prob}} 0
\]

(105)

\[
\text{here} \ \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_n = y) \text{is the underlying} 1 - \alpha \text{quantile function generated by bootstrap algorithm. In particular, prediction quantile generated by algorithm1 satisfies definition3 for sufficiently large repeated times m(see remark2 for an explanation).}
\]

**Proof of lemma7** Define \( c_{1-\alpha}(y) = c_{1-\alpha}(\|X_{n+1} - A Y_n\| Y_n = y) \), \( \hat{c}_{1-\alpha}(y) = c_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_n = y) \), \( \hat{c}_{1-\alpha}^*(y) = \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_n = y) \) and \( \hat{c}_{1-\alpha}(y) = \hat{c}_{1-\alpha}(\|X_{n+1} - \hat{A} Y_n\| Y_n = y) \), then we have

\[
|\hat{c}_{1-\alpha}(y) - \hat{c}_{1-\alpha}(y)| \leq |\hat{c}_{1-\alpha}^*(y) - \hat{c}_{1-\alpha}(y)| + |c_{1-\alpha}(y) - \hat{c}_{1-\alpha}(y)|
\]

(106)

For the second term in (106), from lemma6 and theorem1 we have \( |\hat{c}_{1-\alpha}^*(Y_n) - c_{1-\alpha}^*(Y_n)| \xrightarrow{\text{Prob}} 0 \). For the third term in (106), for any given \( \xi, \epsilon > 0 \), there exists a number \( r > 0 \) such that \( \text{Prob}(\|Y_k\| > r) < \epsilon \) for \( k = \ldots, -1, 0, 1, \ldots \). From lemma5 and equivalence of norm, there exists a constant \( C \) such that \( \| \| \leq C \|_2 \). Notice
that \( \|X_{n+1} - AY_n\| \leq \|X_{n+1} - \tilde{A}Y_n\| + C\|\tilde{A} - A\|_2 \|Y_n\|_2 \) and \( \|X_{n+1} - AY_n\| \geq \|X_{n+1} - \tilde{A}Y_n\| - C\|\tilde{A} - A\|_2 \|Y_n\|_2 \),
if \( \omega \in \Omega \) such that \( c_{1-\alpha}(Y_n) > \tilde{c}_{1-\alpha}(Y_n) + \xi \), we have
\[
\alpha < \text{Prob}(\|X_{n+1} - AY_n\| > \tilde{c}_{1-\alpha}(Y_n) + \xi \|Y_n\|) \leq \text{Prob}(\|X_{n+1} - \tilde{A}Y_n\| > \tilde{c}_{1-\alpha}(Y_n) + \xi - C\|\tilde{A} - A\|_2 \|Y_n\|_2 \|Y_n\|_2)
\] (107)
which is impossible if \( C\|\tilde{A} - A\|_2 \|Y_n\|_2 < \xi \). Thus, \( \text{Prob}(c_{1-\alpha}(Y_n) > \tilde{c}_{1-\alpha}(Y_n) + \xi) \leq \text{Prob}(\|Y_n\|_2 > r) + \text{Prob}(C\|\tilde{A} - A\|_2 \geq \xi) \), which is less than \( 2\epsilon r \) for sufficiently large \( n \). Similarly, if \( c_{1-\alpha}(Y_n) < \tilde{c}_{1-\alpha}(Y_n) - \xi \),
\[
\alpha \geq \text{Prob}(\|X_{n+1} - AY_n\| < \tilde{c}_{1-\alpha}(Y_n) - \xi \|Y_n\|) \geq \text{Prob}(\|X_{n+1} - \tilde{A}Y_n\| < \tilde{c}_{1-\alpha}(Y_n) - \xi + C\|\tilde{A} - A\|_2 \|Y_n\|_2 \|Y_n\|_2)
\] (108)
which is impossible if \( C\|\tilde{A} - A\|_2 \|Y_n\|_2 < \xi \). With similarly discussion we know that \( |c_{1-\alpha}(y) - \tilde{c}_{1-\alpha}(y)| \xrightarrow{n \to \infty} 0 \).

For the first term, for sufficiently large \( n \) with high probability \( \tilde{A} \) and \( \tilde{\Sigma} \) is non-singular, from proof of lemma 6 we know that
\[
\text{Prob}(\|\tilde{A}^* - \tilde{A}\|_2 \geq s|Y_1, \ldots, Y_n) \leq \frac{\tilde{D}}{(n-2)s^2}
\] (109)
with \( \tilde{D} \) be a continuous function of \( \tilde{A} \) and \( \tilde{\Sigma} \) for sufficiently large \( n \) (for distribution of \( \eta^*_k \) coincides with distribution of \( \eta_k \) for any \( k \) and 2th and 4th moment of multivariate normal distribution with mean 0 is continuously related to covariance matrix \( \tilde{\Sigma} \)), thus combine with (107) and (108), from dominated convergence theorem, we let \( T \subset \Omega \) being the set such that \( \forall \omega \in \Omega, \|\tilde{A} - A\|_2,\|\tilde{\Sigma} - \Sigma\|_2 \) being sufficiently small so that \( \|\tilde{A}\| < \|A\| + a < 1, |\tilde{D} - C| < 1 \) with \( C \) being defined in lemma 5 and \( \tilde{A}, \tilde{\Sigma} \) being non-singular. For any \( r = r(n) \),
\[
\lim_{n \to \infty} \text{Prob}(|\tilde{c}_{1-\alpha}(Y_n) - c_{1-\alpha}(Y_n)| > \xi) = \int_\Omega \lim_{n \to \infty} \text{Prob}(|\tilde{c}_{1-\alpha}(Y_n) - c_{1-\alpha}(Y_n)| > \xi|Y_1, \ldots, Y_n) dP
\]
\[
\leq 2 \int_\Omega \lim_{n \to \infty} \left( \text{Prob}(\|Y^*_k\|_2 > r|Y_1, \ldots, Y_n) + \text{Prob}(C\|\tilde{A}^* - \tilde{A}\|_2 \geq \xi|Y_1, \ldots, Y_n) \right) dP + 2 \lim_{n \to \infty} \text{Prob}(T^c)
\]
\[
\leq 2 \int_\Omega \lim_{n \to \infty} \frac{1}{(n-2)s^2} \left( \text{E}(\|X^*_k\|_2|Y_1, \ldots, Y_n) + \text{E}(\|\eta^*_k\|_2|Y_1, \ldots, Y_n) \right) \|\tilde{A}\|_2^2 + 2 \lim_{n \to \infty} \frac{C^2 \|D\|_2^2}{(n-2)s^2} + 2 \lim_{n \to \infty} \frac{2\text{Prob}(T^c)}{n-2)^2}
\] (110)
Notice that \( \text{E}(\|X^*_k\|_2|Y_1, \ldots, Y_n)^2 \leq \text{E}(\|\eta^*_k\|_2^2|Y_1, \ldots, Y_n) = \sum_{i=1}^d \tilde{\lambda}_i^2 \), here \( \lambda_i, i = 1, 2, \ldots, d \) being eigenvalues of \( \tilde{\Sigma} \). Because \( \sum_{i=1}^d \tilde{\lambda}_i^2 \) is a continuous function of matrix \( \Sigma \), so with \( \|\tilde{\Sigma} - \Sigma\|_2 \) being sufficiently small, \( |\sum_{i=1}^d \tilde{\lambda}_i^2 - \sum_{i=1}^d \lambda_i^2| < 1 \) and we choose \( r(n) = (n-2)^{1/3} \), (110) tends to 0. Thus, we prove the result. \( \square \)

Finally, combine with lemma 5 and 6 we can prove the main theorem 3 in this paper.

**Proof of theorem 3** First notice that
\[
\begin{bmatrix}
Y_{n+1} \\
\eta_{n+1}
\end{bmatrix} =
\begin{bmatrix}
A & -A \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
Y_n \\
\eta_n
\end{bmatrix}
+ \begin{bmatrix}
I \\
0
\end{bmatrix}
\begin{bmatrix}
\epsilon_{n+1} \\
\eta_{n+1}
\end{bmatrix}
\] (111)
being an ARMA model and \( A \) being full rank with \( \|A\|_2 < 1 \), so polynomial

\[
f(z) = \det \left( \begin{bmatrix} I & I \\ 0 & I \end{bmatrix} - \begin{bmatrix} A & -A \\ 0 & 0 \end{bmatrix} z \right) = \det(I - Az)
\]  

(112)

has all roots with absolute value being greater than 1. According to theorem 1 in [30], we have \([Y, \eta_i]_T^T\) being mixing process and \( \exists 0 < s < 1 \) with mixing coefficient \( \alpha_{[Y, \eta_i]}(k) = O(s^{|k|}) \) for any \( k \in \mathbb{Z} \). For linear function \( f([Y, \eta_i]_T^T) = Y_i \) is continuous, so \( \sigma - \text{algebra generated by } Y_i, i \leq 0 \) belongs to \( \sigma - \text{algebra generated by } Y_i, i \geq k \) belongs to \( \sigma - \text{algebra generated by } Y_i, i \geq k \) for any \( k \geq 0 \) (like what we do in proof of theorem 2), correspondingly \( \alpha_Y(k) \leq \alpha_{[Y, \eta_i]}(k) = O(s^{|k|}) \) satisfies \([32]\). Notice that \( Y_i = X_i + \eta_i \) for any \( i = 1, 2, ..., n \), from \([6]\) we have

\[
X_{n+1} = AX_n + \epsilon_{n+1} \Rightarrow |\mathcal{F}f_X(t)| = |\mathcal{F}f_X(A^T t)\mathcal{F}f_X(t)| \leq |\mathcal{F}f_X(t)|
\]  

(113)

which implies that \( \mathcal{F}f_X \) satisfies condition A2). According to theorem 2 with bandwidth chosen satisfying \([36]\), especially for bandwidth chosen satisfying condition K2), we have

\[
\|\hat{f}_X - f_X\|_{L_1} \xrightarrow{n \to \infty} 0
\]  

(114)

Define \( \xi_k \) as in \([2]\) for \( k = 1, 2, ..., n - 1 \) and density estimator

\[
\tilde{g}_n(x) = \frac{1}{n - 1} \sum_{k=1}^{n-1} T_n(x - \xi_k), \quad T_n(x) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \exp(-it^T x) \frac{\mathcal{F}K_n(t)}{\mathcal{F}g_n(t)\mathcal{F}g_n(-A^T t)} dt
\]  

(115)

notice that \( \xi_0 \) is independent of \( \xi_i \) for \( i \geq 2 \), mixing coefficient \( \alpha_{\xi_i}(k) = 0 \) for \( |k| \geq 2 \). Thus, \( \xi_i, i = 1, 2, ..., n - 1 \) satisfies condition \([32]\). Notice that

\[
\mathcal{F}f_{X_{n+1} - \Lambda n}(x) = \mathcal{F}f_{\eta_n}(x)\mathcal{F}f_{\eta_n(-A^T x)} \Rightarrow |\mathcal{F}f_{X_{n+1} - \Lambda n}(x)| > 0
\]  

(116)

for any \( x \in \mathbb{R}^d \), so \([62]\) still holds for \( S = \epsilon_1 \). On the other hand, from Cauchy inequality, let \( A = (a_1, ..., a_d) \) with \( a_i \) be column vector for each \( i \), we have, from Cauchy inequality,

\[
x \in \left[ -\frac{a_1}{h_1}, \frac{a_1}{h_1} \right] \times ... \times \left[ -\frac{a_d}{h_d}, \frac{a_d}{h_d} \right] \Rightarrow |a_i^T x| \leq a_i \|x\|_2 \sqrt{\frac{1}{h_1^2} + ... + \frac{1}{h_d^2}}, \quad i = 1, 2, ..., d
\]  

(117)
Thus, let $\Gamma = \left[-\frac{a}{h_1}, \frac{a}{h_1}\right] \times \ldots \times \left[-\frac{a}{h_d}, \frac{a}{h_d}\right]$, $y = -A^T x$,

$$\int_{\Gamma} \frac{1}{|F f_n(x)|^2} \times |F f_n(-A^T x)|^2 \, dx \leq \sqrt{\int_{\Gamma} \frac{1}{|F f_n(x)|^4} \times \int_{\Gamma} \frac{1}{|F f_n(-A^T x)|^4} \, dx}$$

$$\leq \sqrt{\frac{1}{2\pi^d} \prod_{j=1}^{d} \frac{\int_{-a/h_j}^{a/h_j} (x_j^2 + 1)^{2\beta_j} \exp(4\alpha_j |x_j|^\rho_j) \times \frac{1}{2|\det(\Delta)|} \int_{-(A^T)^{-1} y}^{d} \prod_{j=1}^{d} (y_j^2 + 1)^{2\beta_j} \exp(4\alpha_j |y_j|^\rho_j) \, dy}}$$

$$\leq C \prod_{j=1}^{d} h_j^{-4\beta_j - 1 + \rho_j} \exp(4\alpha_j a^\rho_j h_j^{-\rho_j}) \times \frac{1}{2|\det(\Delta)|} \prod_{j=1}^{d} \int_{-a/h_j}^{a/h_j} \frac{1}{\sqrt{\frac{1}{h_j^2} + ... + \frac{1}{h_d^2}}} (y_j^2 + 1)^{2\beta_j} \exp(4\alpha_j |y_j|^\rho_j) \, dy_j$$

$$\leq C (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^{2 \sum_{j=1}^{d} \beta_j + d/2 - \sum_{j=1}^{d} \rho_j / 4} \exp(4 \sum_{j=1}^{d} \alpha_j a^\rho_j (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^{\rho_j/2})$$

$$\leq C (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^{\gamma} \exp(2 \sum_{j=1}^{d} \alpha_j a^\rho_j (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^{\rho_j/2}) \prod_{j=1}^{d} h_j^{-2\beta_j - 1/2 + \rho_j / 2} \exp(2\alpha_j a^\rho_j h_j^{-\rho_j})$$

Let $\gamma = \sum_{j=1}^{d} \beta_j + d/4 - \sum_{j=1}^{d} \rho_j / 4$, from [64] and [68], we have

$$E\|\hat{g}_n - f_i\|^2 \leq C \sum_{j=1}^{d} h_j^{2\beta_j} \exp(-2\alpha_j h_j^{-\rho_j})$$

$$+ \frac{C}{n} (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^\gamma \exp(2 \sum_{j=1}^{d} \alpha_j a^\rho_j (\frac{1}{h_1^2} + ... + \frac{1}{h_d^2})^{\rho_j/2}) \prod_{j=1}^{d} h_j^{-2\beta_j - 1/2 + \rho_j / 2} \exp(2\alpha_j a^\rho_j h_j^{-\rho_j})$$

which converges to 0 if bandwidth $h(n)$ satisfies K2).

From assumption A3) and theorem 8.22 in [17], we know that $\text{Re} F f_n$, $\text{Im} F f_n$ are differentiable. Let $\hat{\xi}_k$ being defined in [8] for $i = 1, 2, ..., n - 1$, let $\nabla$ being gradient operator,

$$\|\hat{g}_n - f_i\|^2 \leq 2\|\hat{g}_n - f_i\|^2 + \frac{4}{(2\pi)^{2n}} \sum_{k=1}^{n} \int_{\mathbb{R}^d} \frac{F K_n(x)}{|F f_n(x)|^2} \frac{F K_n(x)}{|F f_n(-A^T x)|^2} \exp(i\nabla \cdot \hat{\xi}_k) \exp(-i\nabla \cdot \hat{\xi}_k) \, dx$$

$$\leq 2\|\hat{g}_n - f_i\|^2 + \frac{4}{(2\pi)^{2n}} \sum_{k=1}^{n} \int_{\mathbb{R}^d} \frac{F K_n(x)}{|F f_n(x)|^2} \frac{F f_n(-A^T x) - f_n(-A^T x)^2}{|F f_n(-A^T x)|^2} \, dx$$

$$+ \frac{4}{(2\pi)^{2n}} \sum_{k=1}^{n} \int_{\mathbb{R}^d} \frac{F K_n(x)}{|F f_n(x)|^2} \frac{F f_n(-A^T x) - f_n(-A^T x)^2}{|F f_n(-A^T x)|^2} \, dx$$

$$= 2\|\hat{g}_n - f_i\|^2 + \frac{4}{(2\pi)^{2n}} \sum_{k=1}^{n} \int_{\mathbb{R}^d} \frac{F K_n(x)}{|F f_n(x)|^2} \frac{|\nabla \text{Re} F f_n(\hat{\xi}_k) + i\nabla \text{Im} F f_n(\hat{\xi}_k) |^2 (A^T - \hat{A}) Y_k}{|\nabla \text{Im} F f_n(\hat{\xi}_k)|^2 (A^T - \hat{A}) Y_k} \, dx$$

$$+ \frac{4}{(2\pi)^{2n}} \int_{\mathbb{R}^d} \frac{F K_n(x)}{|F f_n(x)|^2} \frac{|\nabla \text{Re} F f_n(\hat{\xi}_k) + i\nabla \text{Im} F f_n(\hat{\xi}_k) |^2 (A^T - \hat{A}) Y_k}{|\nabla \text{Im} F f_n(\hat{\xi}_k)|^2 (A^T - \hat{A}) Y_k} \, dx$$

31
Thus, for sufficiently large $n$, from \( \{ \text{some constant} C \} \), we know that $x$ satisfies $A2$). From condition $K2$ and (117), we have for any $n > s > \frac{1}{4}$ for $x > 0$, $n$ for any $x$, we have

$$
\mathbb{P} \left( \frac{1}{n} \sum_{k=1}^{n} \left\| Y_k \right\|_2^2 > r^2 \right) + \mathbb{P} \left( \left\| \hat{A} - A \right\|_2 > s \right) + \mathbb{P} \left( \frac{8}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{f_{K_n}(x)}{f_{\hat{f}_n}(x)} |x|^2 \times \frac{1}{2} \left\| x \right\|^2 s^2 < \frac{1}{4} \right)
$$

From (76), we know that $E(\frac{1}{n} \sum_{k=1}^{n} \left\| Y_k \right\|_2^2) \leq E\left\| Y_k \right\|_2^2 \leq 8E\left\| X_k \right\|_2^2 + 8E\left\| Y_k \right\|_2^2 < \infty$. If $r > E\left\| f_{K_n}(x) / f_{\hat{f}_n}(x) - A^T x \right\|^2 \times \frac{1}{4} \left\| x \right\|^2 s^2$ is of order $o(1)$, then the third term in (123) is 0 for sufficiently large $n$. From Cauchy inequality, for some constant $C$,

$$
\sum_{j=1}^{d} \int_{\mathbb{R}^d} \frac{f_{K_n}(x)}{f_{\hat{f}_n}(x)} |x|^2 \leq \left\| f_{K_n} \right\|^2_{L_\infty} \sum_{j=1}^{d} \left[ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{1}{|x|^4} \left| f_{K_n}(x) / f_{\hat{f}_n}(x) - A^T x \right|^2 \right]
$$

$$
\leq C \sum_{j=1}^{d} h_j^{-2} \left\{ \sum_{k=1}^{d} h_k^{-2\beta_k - 1/2 + \rho_k/2} \exp(2\alpha_k a^{:\gamma_k \rho_k}) \left( \frac{1}{h_1} + \ldots + \frac{1}{h_d} \right)^\gamma \exp \left( 2 \sum_{j=1}^{d} \alpha_j a^{:\rho_j} \left( \frac{1}{h_1} + \ldots + \frac{1}{h_d} \right)^\rho_j \right) \right\}
$$

As long as

$$
r_j = \frac{1}{n} \sum_{k=1}^{d} h_k^{-2} \left\{ \sum_{j=1}^{d} h_k^{-2\beta_k - 1/2 + \rho_k/2} \exp(2\alpha_k a^{:\gamma_k \rho_k}) \left( \frac{1}{h_1} + \ldots + \frac{1}{h_d} \right)^\gamma \exp \left( 2 \sum_{j=1}^{d} \alpha_j a^{:\rho_j} \left( \frac{1}{h_1} + \ldots + \frac{1}{h_d} \right)^\rho_j \right) \right\}
$$

for $j = 1, 2, \ldots, d$, especially when $K2$ holds, (123) tends to 0 for suitable $s, t$ (for example, let $s = \frac{1}{\sqrt{\sum_{j=1}^{d} r_j^{1/2}}} \to 0$ and $t = \frac{1}{(\sum_{j=1}^{d} r_j^{1/2})^{1/2}} \to \infty$).

For the third term in (122), from condition $A3$ and theorem 8.22 in [17], $||\nabla \Re Ff(\zeta)||_2 + ||\nabla \Im Ff(\zeta)||_2$ for any $\zeta_1, \zeta_2$ is bounded by a constant $D$. Notice that $(1 + x^2)^{-\beta_j/2}$ and $\exp(-\alpha_j |x|^\rho_j)$ are decreasing for parameters satisfying $A2$, and from condition $K2$ and (117), we have for any $x \neq 0 \in [-\frac{a}{h_1}, \frac{a}{h_1}] \times \ldots \times [-\frac{a}{h_d}, \frac{a}{h_d}]$,

$$
\frac{|F_{f_0}(A^T x)|}{\|x\|_2} \geq c_{n} \sum_{j=1}^{d} \beta_j (\sum_{j=1}^{d} 1 + \frac{\sum_{j=1}^{d} 1}{h_j^{\beta_j}})^{-\beta_j/2} \times \exp(-\sum_{j=1}^{d} \alpha_j a^{\rho_j} (\sum_{j=1}^{d} 1)^{\rho_j/2}) \sqrt{a^{2}/h_1^{2} + \ldots + a^{2}/h_d^{2}}
$$

$$
\geq \frac{c_n \min(h_1, \ldots, h_d)}{2a^{1+\sum_{j=1}^{d} \beta_j}} \sum_{j=1}^{d} \frac{1}{h_j^{\beta_j}} \times \exp(-\sum_{j=1}^{d} \alpha_j a^{\rho_j} (\sum_{j=1}^{d} 1)^{\rho_j/2}) \geq \frac{1}{n^{1/2}}
$$

Thus, for sufficiently large $n$ with high probability $|F_{f_0}(A^T x)| > 2D \|A^T - A^T\|_2 \|x\|_2 > 0$. If it is satisfied,
let $\gamma_1 = 2 \sum_{j=1}^{d} \beta_j + d/4 - \sum_{j=1}^{d} \beta_j/4$, there exists a constant $C$ such that

$$\int_{\mathbb{R}^d} \left| \langle \nabla \text{Re} F \phi, \nabla \text{Im} F \phi \rangle \right|^2 (A^T - \tilde{A}^T) |x|^2 \leq \sum_{j=1}^{d} \int_{\mathbb{R}^d} \left| F K_h(x) \right|^2 D^2 \left\| A^T - \tilde{A}^T \right\|_2^2 \left\| x_j \right\|^2$$

$$\leq 4D^2 \left\| A^T - \tilde{A}^T \right\|_2^2 \left\| F K \right\|_2^2 \sum_{j=1}^{d} \left( \int_{\mathbb{R}^d} \left| \nabla \text{Re} F \phi \right|^2 \left| \nabla \text{Im} F \phi \right|^2 (A^T - \tilde{A}^T) |x|^2 \right)^{\gamma_j} \exp(4 \sum_{j=1}^{d} a_{j}^{\beta_j} \left( \frac{1}{h_j^2} + \frac{1}{h_j^2} \right)^{\beta_j/2})$$

(127)

correspondingly, if K2) is valid, for any $\epsilon > 0$, let $\Gamma = [-\frac{a}{h^2}, \frac{a}{h^2}] \times \ldots \times [-\frac{a}{h^2}, \frac{a}{h^2}] - \{0\}$,

$$\text{Prob} \left( \frac{1}{n} \sum_{k=1}^{n} \left( \frac{1}{n} \right)^2 \left| \int_{\mathbb{R}^d} \left| F K_h(x) \right|^2 \left| \nabla \text{Re} F \phi \right|^2 \left| \nabla \text{Im} F \phi \right|^2 (A^T - \tilde{A}^T) |x|^2 \right)^{\gamma_j} \exp(4 \sum_{j=1}^{d} a_{j}^{\beta_j} \left( \frac{1}{h_j^2} + \frac{1}{h_j^2} \right)^{\beta_j/2}) > \epsilon \right)$$

$$\leq \text{Prob}(\exists x \in \Gamma, |F f(x) - A^T x| \leq 2D \left\| A^T - \tilde{A}^T \right\|_2 |x|^2)$$

$$+ \text{Prob}(C \left\| A^T - \tilde{A}^T \right\|_2 \sum_{j=1}^{d} h_j^{-2} \left\| h_j^{-2} \right\|_2 \left\| A^T - \tilde{A}^T \right\|_2 |x_j|^2 \exp(2a_{j}^{\beta_j} h_j^{-\beta_j}) \right)$$

$$\times \left( \frac{1}{h_j^2} + \frac{1}{h_j^2} \right)^{\gamma_j} \exp(4 \sum_{j=1}^{d} a_{j}^{\beta_j} \left( \frac{1}{h_j^2} + \frac{1}{h_j^2} \right)^{\beta_j/2}) > \epsilon \right) \rightarrow 0$$

(128)

Similar with (71) to (74), if K2) holds, we have for any $\epsilon > 0$ and sufficiently large $n$,

$$\text{Prob}(\max_{\Omega} (\text{Re} \tilde{f}_x, 0) \sum_{j=1}^{d} |_{-s_j, s_j} - f, \Omega \epsilon) \leq \text{Prob}(2^{d+1} s_1 \ldots s_d \int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} | \tilde{f}_x - f, \Omega \epsilon |^2 + 2(\int_{[-s_1, s_1] \times \ldots \times [-s_d, s_d]} f) \epsilon^2$$

$$\text{Prob}(\max_{\Omega} (\text{Re} \tilde{f}_x, 0) - 1) \leq \text{Prob}(\max_{\Omega} (\text{Re} \tilde{f}_x, 0) - f) + \text{Prob}(\max_{\Omega} (\text{Re} \tilde{f}_x, 0) - f) \epsilon) \rightarrow 0$$

(130)

which implies that $\| \tilde{f}_x - f \|_{L_1} \leq \frac{\epsilon}{n^{\text{prob}}} \rightarrow 0$. Finally, from Young’s inequality,

$$\| f \|_{L_\infty} \leq \| f \|_{L_1} \leq \| f \|_{L_\infty} \leq \| f \|_{L_1}$$

(131)

and from condition A2), $\| f \|_{L_\infty} < \infty$, combine with (114) we prove 1).

From definition 1 and 2, $\tilde{f}_x \geq 0$ for any $x \in \mathbb{R}^d$ and $\int_{\mathbb{R}^d} \tilde{f}_x \neq 1$ if and only if

$$\int_{-s_1}^{s_1} \ldots \int_{-s_d}^{s_d} \max(\text{Re} \tilde{f}_x(x_1, \ldots, x_d), 0) dx_1 \ldots dx_d = 0$$

(132)

which has negligible probability as $n \rightarrow \infty$ according to (130), and this also holds for $\tilde{f}_x$ (according to (72)).

According to lemma 6, we know that 2) is valid. 3) is a direct result from lemma 7.

33
References

[1] 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.

[2] Dimitris N. Politis. *Model-Free Prediction and Regression*. Springer, Cham, 1st edition, 2015.

[3] George A.F. Seber and Alan J. Lee. *Linear Regression Analysis*. John Wiley & Sons, 2002.

[4] Robert A. Stine. Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*, 80(392):1026–1031, 1985.

[5] Lori A. Thombs and William R. Schucany. Bootstrap prediction intervals for autoregression. *Journal of the American Statistical Association*, 85(410):486–492, 1990.

[6] Michael Wolf and Dan Wunderli. Bootstrap joint prediction regions. *Journal of Time Series Analysis*, 36(3):352–376.

[7] Li Pan and Dimitris N. Politis. Bootstrap prediction intervals for markov processes. *Computational Statistics & Data Analysis*, 100:467 – 494, 2016.

[8] Donald E. Catlin. *Estimation, Control, and the Discrete Kalman Filter*. Springer, New York, NY, 1989.

[9] Andrew C. Harvey. *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, 1990.

[10] Peter J. Brockwell and Richard A. Davis. *Time Series: Theory and Methods*. Springer-Verlag New York, 2nd edition, 1991.

[11] PIET DE JONG. The likelihood for a state space model. *Biometrika*, 75(1):165–169, 03 1988.

[12] David S. Stoffer and Kent D. Wall. Bootstrapping state-space models: Gaussian maximum likelihood estimation and the kalman filter. *Journal of the American Statistical Association*, 86(416):1024–1033, 1991.

[13] William N. Anderson, George B. Kleindorfer, Paul R. Kleindorfer, and Michael B. Woodroofe. Consistent estimates of the parameters of a linear system. *The Annals of Mathematical Statistics*, 40(6):2064–2075, 1969.

[14] Craig F. Ansley and Robert Kohn. Estimation, filtering, and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics*, 13(4):1286–1316, 1985.

[15] Genshiro Kitagawa. Non-gaussian state-space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82(400):1032–1041, 1987.

[16] F. Comte and C. Lacour. Anisotropic adaptive kernel deconvolution. *Annales de l’I.H.P. Probabilités et statistiques*, 49(2):569–609, 2013.

[17] Gerald B Folland. *Real analysis: modern techniques and their applications; 2nd ed*. Pure and Applied Mathematics. Wiley, Hoboken, NJ, 1999.

[18] G. Golub and W. Kahan. Calculating the singular values and pseudo-inverse of a matrix. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 2(2):205–224, 1965.
[19] Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer-Verlag New York, 1999.

[20] Erhan Cinlar. *Probability and Stochastics*. Springer-Verlag New York, 2011.

[21] Jianqing Fan. Global behavior of deconvolution kernel estimates. *Statistica Sinica*, 1(2):541–551, 1991.

[22] Loukas Grafakos. *Classical Fourier Analysis*. Springer-Verlag New York, 2014.

[23] Nicholas J. Higham. Computing real square roots of a real matrix. *Linear Algebra and its Applications*, 88-89:405 – 430, 1987.

[24] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1998.

[25] James E. Gentle. *Computational Statistics*. Springer-Verlag New York, 1st edition, 2009.

[26] Bert Van Es, Peter Spreij, and Harry Van Zanten. Nonparametric volatility density estimation for discrete time models. *Journal of Nonparametric Statistics*, 17(2):237–249, 2005.

[27] Henrik Spliid. A fast estimation method for the vector autoregressive moving average model with exogenous variables. *Journal of the American Statistical Association*, 78(384):843–849, 1983.

[28] Paul Doukhan. *Mixing*. Springer-Verlag New York, 1994.

[29] Roger A. Horn and Charles R. Johnson. *Matrix Analysis*. Cambridge University Press, New York, NY, USA, 2nd edition, 2012.

[30] Abdelkader Mokkadem. Mixing properties of arma processes. *Stochastic Processes and their Applications*, 29(2):309 – 315, 1988.