Negative Moment Bounds for Stochastic Regression Models with Deterministic Trends and Their Applications to Prediction Problems

Chien-Ming Chi
University of Southern California

Ching-Kang Ing
National Tsing Hua University

Shu-Hui Yu
National University of Kaohsiung

November 3, 2020

Abstract

We establish negative moment bounds for the minimum eigenvalue of the normalized Fisher information matrix in a stochastic regression model with a deterministic time trend. This result enables us to develop an asymptotic expression for the mean squared prediction error (MSPE) of the least squares predictor in the aforementioned model. Our asymptotic expression not only helps better understand how the MSPE is affected by the deterministic and random components, but also
inspires an intriguing proof of the formula for the sum of elements in the inverse of
the Cauchy/Hilbert matrix from a prediction perspective.

Keywords: Cauchy matrix, Hilbert matrix, mean squared prediction error, mini-
mum eigenvalue, negative moment bound, stochastic regression model

1 Introduction

The stochastic regression model is one of the most widely used statistical models due to its
broad applications in engineering, economics, medicine, and many other scientific fields.
In their seminal paper, Lai and Wei (1982) laid the theoretical foundations of parameter
estimation in such a model. In particular, they proposed a set of weakest possible condi-
tions under which the linear least squares estimate achieves strong consistency. Lai and
Wei’s (1982) paper inspired a great deal of exciting work, bringing insights into prediction,
model selection, non-linear estimation, stochastic approximation, and adaptive control;
see, e.g., Chen and Guo (1986), Lai and Wei (1986), Wei (1987, 1992), Lai (1994), Lai
and Lee (1997), Chen et al. (1999), and Gerencsér et al. (2009).

One of the most important purposes of statistical modeling is to predict future values.
The performance of a prediction method is usually evaluated by two different measures:
the accumulated prediction error (APE) and the mean squared prediction error (MSPE).
Model selection based on these two types of errors also attracted a lot of attention from
researchers and practitioners. Wei (1987) provided asymptotic expressions for the APEs
of the least squares predictors in stochastic regression models. Model selection based on
the APE was explored by Rissanen (1986), Wax (1988), Hannan et al. (1989), Hemerly
and Davis (1989), Wei (1992), Speed and Yu (1994), West (1996), Lai and Lee (1997), Ing (2004, 2007), Ing et al. (2009), and Ing and Yang (2014). Asymptotic expressions for the MSPEs of least squares predictors were derived in a variety of time series models; see, e.g., Fuller and Hasza (1981), Kunitomo and Yamamoto (1985), Gerencsér (1992), Ing (2003), Ing et al. (2009), Chan and Ing (2011), and Chan et al. (2013). There are also quite a few model selection methods proposed based on minimizing the MSPE; see Shibata (1980), Bhansali (1996), Lee and Karagrigoriou (2001), Ing and Wei (2005), Ing et al. (2012), and Hsu et al. (2019).

Since many time series data exhibit polynomial or other deterministic time trends, parameter estimation and hypothesis testing in time series models with drifts have been considered by a number of authors; see, e.g., Chan (1989), Hamilton (1994), and Stock (1994). On the other hand, most existing studies on the MSPE have focused on the case where the underlying time series model has a constant mean. Although Ing (2003) derived an asymptotic expression for the MSPE of the least squares predictor in an autoregressive (AR) model around a polynomial trend, it seems difficult to apply his result to more general time series models. In addition, his derivation heavily relied on a negative moment bound for the minimum eigenvalue of the normalized Fisher information matrix of a non-constant mean, whose proof, however, was not rigorously given. In this paper, we fill this gap by investigating the MSPEs of the least squares predictors in autoregressive exogenous (ARX) models (an important class of stochastic regression models) with deterministic trends satisfying general conditions. We first establish negative moment bounds for the minimum eigenvalue of the normalized Fisher information matrix, \( \hat{R}_n \), associated with
this model in a rigorous manner. With the help of these bounds, we provide an asymptotic expression for the MSPE of the least squares predictor, which is the sum of two terms accounting for the variations due to estimating the time trend and ARX components, respectively. This result helps us better understand how the MSPE is affected by the model’s deterministic and random components.

Our asymptotic expression shows that the MSPE due to estimating the polynomial time trend is related to the sum of elements in the inverse of the Hilbert matrix, which, in turn, is a special case of the symmetric Cauchy matrix. The formula for the sum of elements in the latter matrix’s inverse was given by Schechter (1959) through Lagrange’s interpolation method. The connection between the MSPE and Cauchy/Hilbert matrix triggers us to ask if there is an alternative proof of the formula from a prediction perspective. By establishing an intriguing link between the MSPE and APE, we show that the answer to this question is affirmative.

The rest of the paper is organized as follows. In Section 2, we establish negative moment bounds for the minimum eigenvalue of a matrix associated with $\hat{R}_n$. In Section 3.1, asymptotic expressions for the MSPEs of the least squares predictors in ARX models with general time trends are given. We further illustrate the results using the polynomial and periodic time trends. In Section 3.2, a statistical proof of the formula for the sum of elements in the inverse of the Cauchy matrix is provided. We conclude in Section 4. All proofs of the theorems in Sections 2 and 3.1 and other technical details are relegated to the Appendix.
2 Negative Moment Bounds

Let $k$ and $m$ be positive integers. We start by considering a $km$-dimensional time series,

$$Y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t,j},$$

(2.1)

where $\varepsilon_{t,j} = (\delta_{t-jk}, \ldots, \delta_{t-(j+1)k+1})^T$, $\{\delta_t\}$ is a sequence of $m$-dimensional independent random vectors satisfying $E(\delta_t) = 0$ and $E(\delta_t \delta_t^T) = \Sigma > 0$, $C_j$s are $km \times km$ coefficient matrices, $C_0$ is invertible, and

$$\sum_{j=0}^{\infty} \|C_j\|_F^2 < \infty.$$  (2.2)

Here, $\|A\|_F$ denotes the Frobenius norm of matrix $A$. Many time series regression models have explanatory vectors satisfying (2.1). Here, we give two examples.

Example 2.1 Let $z_t = \sum_{j=0}^{\infty} D_j \delta_{t-j}$ be an $m$-dimensional stationary time series, where $\sum_{j=0}^{\infty} \|D_j\|_F^2 < \infty$, $D_0$ is invertible, and $\{\delta_t\}$ is defined as in model (2.1). Then,

$$
\begin{pmatrix}
z_t \\
\vdots \\
z_{t-k+1}
\end{pmatrix} = \sum_{j=0}^{\infty} 
\begin{pmatrix}
D_{jk} & \cdots & D_{j(j+1)k-1} \\
\vdots & \ddots & \vdots \\
D_{(j-1)k+1} & \cdots & D_{jk}
\end{pmatrix}
\begin{pmatrix}
\delta_{t-jk} \\
\vdots \\
\delta_{t-(j+1)k+1}
\end{pmatrix},
$$

(2.3)
where $D_l = 0$ if $l < 0$. Hence (2.1) holds with $Y_t = (z_t^T, \cdots, z_{t-k+1}^T)^T$,

$$C_j = \begin{pmatrix}
D_jk & \cdots & D_{(j+1)k-1} \\
\vdots & \ddots & \vdots \\
D_{(j-1)k+1} & \cdots & D_{jk}
\end{pmatrix},$$

and $\varepsilon_t = (\delta_t^T, \cdots, \delta_t^{T(j+1)k-1})^T$. One may use the following vector AR model to make prediction,

$$z_{t+1} = \sum_{j=1}^{k} \Theta_j z_{t+1-j} + \epsilon_{t+1},$$

where $\Theta_j$ are $m \times m$ coefficient matrices and $\epsilon_{t+1}$ is the model error, which can be serially correlated if (2.4) is misspecified. It is clear that the explanatory vector of model (2.4) is given on the left-hand side of (2.3).

Example 2.2 Consider an autoregressive exogenous (ARX) model,

$$v_t = \sum_{j=1}^{k_0} a_j v_{t-j} + \sum_{l=1}^{d} \sum_{j=1}^{k_l} \theta_j(l) z_{t-j-l} + \varepsilon_t,$$

where $d, k_0, \ldots, k_d$ are positive integers, $a_j$ and $\theta_j(l)$ are unknown coefficients,

$$1 - a_1 z - \cdots - a_{k_0} z^{k_0} \neq 0, \ |z| \leq 1,$$  

$$(z_{t-l}, \cdots, z_{t-k_l+1})^T, l = 1, \ldots, d,$$ are exogenous variables admitting the following
MA(∞) representation,

$$z_t(l) = \sum_{j=0}^{\infty} b_j(l) \varepsilon_{t-j}(l), \quad (2.7)$$

with $b_0(l) = 1$ and $\sum_{j=0}^{\infty} b_j^2(l) < \infty$, and $\delta_t = (\varepsilon_t, \varepsilon_t(1), \ldots, \varepsilon_t(d))^\top$, $t = 1, \ldots, n$, are independent noises satisfying $E(\delta_t) = 0$ and $(\sigma_{ij})_{1 \leq i, j \leq d+1} = E(\delta_i \delta_j^\top) > 0$. By (2.6) and (2.7), there exist $\eta_j$, $j \geq 0$, with $\eta_0 = (1, 0, \ldots, 0)^\top$ and $\sum_{j=1}^{\infty} \|\eta_j\|^2 < \infty$, such that

$$v_t = \sum_{j=0}^{\infty} \eta_j^\top \delta_t - j, \quad (2.8)$$

noting that $\| \cdot \|$ denotes the Euclidean norm. Let $\bar{k} = \max\{k_0, \ldots, k_d\}$. Then

$$Y_t = (v_t, z_t(1), \ldots, z_t(d), \ldots, v_{t-\bar{k}+1}, z_{t-\bar{k}+1}(1), \ldots, z_{t-\bar{k}+1}(d))^\top \quad (2.9)$$

can be viewed as an explanatory vector of model (2.5) containing possibly redundant components. It follows from (2.7) and (2.8) that

$$Y_t = \sum_{j=0}^{\infty} C_j \varepsilon_{t,j},$$
where \( \varepsilon_{t,j} = (\delta_{t-j\tilde{k}}^T, \ldots, \delta_{t-(j+1)\tilde{k}+1}^T)^T \) and

\[
C_j = \begin{pmatrix}
C_{j,11} & \cdots & C_{j,1\tilde{k}} \\
\vdots & \ddots & \vdots \\
C_{j,\tilde{k}1} & \cdots & C_{j,\tilde{k}\tilde{k}}
\end{pmatrix},
\]

in which

\[
C_{j,tt} = \begin{pmatrix}
\eta_{j\tilde{k}-t+1,1} & \eta_{j\tilde{k}-t+1,2} & \cdots & \eta_{j\tilde{k}-t+1,\tilde{k}+1} \\
0 & b_{j\tilde{k}-t+1(1)} & 0 & \cdots & 0 \\
\vdots & 0 & b_{j\tilde{k}-t+1(2)} & 0 & \vdots \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & 0 & b_{j\tilde{k}-t+1(d)}
\end{pmatrix},
\]

is a \((d+1) \times (d+1)\) matrix with \((\eta_{j\tilde{k}-t+1,1}, \ldots, \eta_{j\tilde{k}-t+1,\tilde{k}+1})^T = \eta_{j\tilde{k}-t+1}\) and \(\eta_{h,t1} = b_h(l_2) = 0\) if \(h < 0\). Since \(\sum_{j=0}^{\infty} \|C_j\|_F^2 < \infty\) and \(C_0\) is invertible due to \(C_{0,tt} = I_{d+1}\) (the \((d+1)\)-dimensional identity matrix) and \(C_{0,tt} = 0\) if \(t > l\), we conclude that the \(Y_t\) in (2.9) fulfills (2.1) with \(k = \tilde{k}\) and \(m = d + 1\).

Assuming that (2.1) holds and there exist \(\delta, M, \alpha > 0\) such that for any \(0 < w - u \leq \delta\),

\[
\sup_{-\infty < t < \infty} \sup_{||\nu||=1} \mathbb{P}(u < \nu^T \delta_t \leq w) \leq M(w - u)^\alpha, \tag{2.10}
\]

Findley and Wei (2002) showed that for any \(q \geq 1\),

\[
\mathbb{E} \left( \lambda_{\min}^{-q} (n^{-1} \sum_{t=1}^{n} Y_t Y_t^\top) \right) = O(1), \tag{2.11}
\]
where \( n \) is the sample size and \( \lambda_{\text{min}}(A) \) denotes the minimum eigenvalue of matrix \( A \). With the help of (2.11), they presented the first mathematically complete derivation of an analogous property of AIC for comparing vector autoregressions fit to weakly stationary series. Using (2.11) and an argument in Findley and Wei (2002), one can also obtain an asymptotic expression for the MSPE of the least squares predictors in model (2.4) (or (2.5)) in terms of the sample size, the variance of the model error, and the number of the estimated parameters.

When a deterministic trend (containing \( p \) variables with \( p \geq 1 \)) is taken into account, a natural generalization of (2.11) is

\[
E\left(\lambda_{\text{min}}^{-\frac{q}{2}}(n^{-1} \sum_{t=1}^{n} \omega_t \omega_t^\top)\right) = O(1),
\]

(2.12)

where \( \omega_t = (x_t^{(n)} \top, Y_t \top) \top \) and \( x_t^{(n)} \in \mathbb{R}^p \), possibly depending on \( n \), is the normalized time trend variables satisfying

\[
\sup_{1 \leq t \leq n} \|x_t^{(n)}\| < M_1,
\]

(2.13)

for some positive constant \( M_1 \). One would expect that (2.12) holds under the additional assumption,

\[
\lim \inf_{n \to \infty} \lambda_{\text{min}}(n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)} \top) > 0,
\]

(2.14)

which is commonly made on the fixed design matrix. The proof of (2.12), however, is far
from being trivial. The main reason is that the proof of (2.11) is built on the property that for any \( a \in \mathbb{R}^{km} \) with \( \|a\| = 1 \), the conditional distribution of \((a^\top Y_t)^2\) given information up to time \( t - l \) is sufficiently smooth at the origin as long as \( l \) is large enough. This property is ensured by (2.10), but is no longer valid when \( Y_t \) is replaced by \( \omega_t \). With the appearance of \( x_t^{(n)} \), it is easy to find a unit vector \( a \in \mathbb{R}^{km+p} \) such that \( a^\top \omega_t = 0 \). In Lemma A.2, we provide a characterization of (2.14). This characterization is not only of independent interest, but it also inspires a proof strategy to bypass the above difficulty.

The main result of this section is given in the following theorem.

**Theorem 2.1** Assume (2.1), (2.2), (2.10), (2.13), (2.14), and

\[
\sup_{-\infty < t < \infty} \max_{1 \leq i \leq m} \mathbb{E}|\delta_{t,i}|^{2q} < \infty, \tag{2.15}
\]

where \( \gamma > 1 \) and \((\delta_{t,1}, \ldots, \delta_{t,m})^\top = \delta_t\). Then, for \( 0 < q < \gamma \), (2.12) follows.

## 3 Applications

### 3.1 Mean squared prediction errors

In this section, we focus on the ARX model around a deterministic time trend,

\[
y_t = \sum_{j=1}^{p} \beta_j s_{t,j} + \sum_{j=1}^{k_0} a_j y_{t-j} + \sum_{l=1}^{d} \sum_{j=1}^{k_l} \theta_j(l) z_{t-j}(l) + \varepsilon_t \tag{3.1}
\]
where \( p, d, \) and \( k_0, \ldots, k_d \) are positive integers, \( \beta_j, a_j, \) and \( \theta_j(l) \) are unknown coefficients, with \( a_j \) satisfying (2.6), \( z_{t-1}(l) = (z_{t-1}(l), \ldots, z_{t-k_l}(l))^\top, 1 \leq l \leq d, \) are exogenous variables admitting the \( \text{MA}(\infty) \) representations described in (2.7), \( s_t = (s_{t,1}, \ldots, s_{t,p})^\top \) are deterministic variables, and \( \delta_t = (\varepsilon_t, \varepsilon_t(1), \ldots, \varepsilon_t(d))^\top \) are the same as the one in Example 2.2. Let \( P_t = (s_t^\top, y_{t-1}^\top, z_{t-1}^\top(1), \ldots, z_{t-1}^\top(d))^\top, \) where \( y_t = (y_t, \ldots, y_{t-k_0+1})^\top. \) Having observed \( y_1, \ldots, y_n \) and \( P_1, \ldots, P_{n+1}, \) we are interested in predicting \( y_{n+1} \) using the least squares predictor,

\[
\hat{y}_{n+1} = P_{n+1}^\top \left( \sum_{t=1}^{n} P_t P_t^\top \right)^{-1} \sum_{t=1}^{n} P_t y_t,
\]

provided the inverse of \( \sum_{t=1}^{n} P_t P_t^\top \) exists.

To analyze the MSPE, \( \mathbb{E}(y_{n+1} - \hat{y}_{n+1})^2, \) of \( \hat{y}_{n+1}, \) we impose the following conditions on the deterministic terms \( s_t: \) there exists a \( p \times p \) matrix \( D \) such that for any \( t, \)

\[
s_{t-1} = D s_t,
\]

and

\[
I_p - \sum_{j=1}^{k_0} a_j D^j \text{ is invertible.}
\]

By (3.3) and (3.4), it can be shown that

\[
y_t = \beta^\top s_t + v_t,
\]
where \( v_t \) is defined in (2.8) and \( \beta^* = \beta^T(I_p - \sum_{j=1}^{k_0} a_j D_j)^{-1} \) with \( \beta = (\beta_1, \ldots, \beta_p)^T \).

Many commonly used deterministic trends fulfill (3.3) and (3.4). For example, in the case of the polynomial trend,

\[
s_t = (1, t, \ldots, t^{p-1})^T, \quad p \geq 1, \tag{3.6}
\]

we have \( D = (D_{ij})_{1 \leq i, j \leq p} \), where \( D_{ij} = 0 \) if \( 1 \leq i < j \leq p \), and \( C_{i-j}^{i-1}(-1)^{i-j} \) if \( 1 \leq j < i \leq p \), where \( C_{i-j}^{i-1} = (i-1)!/(i-j)!(j-1)! \). Since \( D_j, j \geq 1 \), are lower triangular matrices with diagonal entries 1, (3.4) holds when (2.6) is assumed. For the periodic trend,

\[
s_t = (1, \sin \nu_1 t, \cos \nu_1 t \ldots, \sin \nu_h t, \cos \nu_h t)^T, \tag{3.7}
\]

where \( h \geq 1 \) and \( 0 < \nu_1 < \cdots < \nu_h < \pi \), we have \( D = \text{Diag}(1, \nu_1, \ldots, \nu_h) \), where

\[
\nu_j = \begin{pmatrix} \cos \nu_j & -\sin \nu_j \\ \sin \nu_j & \cos \nu_j \end{pmatrix}.
\]

Also, (3.4) follows from (2.6). By (3.3) and (3.5), the trend in \( y_{t-1} \) can be removed through a linear transformation of \( P_t \),

\[
y_{t-1} - G s_t = v_{t-1} = (v_{t-1}, \ldots, v_{t-k_0})^T, \tag{3.8}
\]

where \( G^T = (D^T \beta^* \cdots D^k \beta^*) \). Suppose that there exists a \( p \times p \) non-random matrix \( Q_n \) such that (2.13) and (2.14) hold with \( x_t^{(n)} = Q_n s_t \). Then, this assumption and (3.8)
together suggest a linear transformation, \( F_n \), of \( P_t \) that depends on \( G \) and \( Q_n \) and satisfies

\[
F_n P_t = (x_t^{(n)}\top, v_{t-1}^\top, z_{t-1}^\top(1), \ldots, z_{t-1}^\top(d))\top,
\]

in which each component has the same order of magnitude and the deterministic and random components are completely separated. Define \( G_t^{(n)} = F_n P_t \) and \( \hat{R}_n = n^{-1} \sum_{t=1}^n G_t^{(n)} G_t^{(n)\top} \).

Since \((v_{t-1}^\top, z_{t-1}^\top(1), \ldots, z_{t-1}^\top(d))^\top\) is a subvector of \( Y_{t-1} \) defined in (2.9), it follows from Theorem 2.1 that for \( 0 < q < \gamma \),

\[
E\left( \lambda_{\min}^{-q}(\hat{R}_n) \right) = O(1),
\]

provided (2.10) holds with \( m = d + 1 \) and

\[
\sup_{-\infty < t < \infty} E|\varepsilon_t|^{2\gamma} + \sup_{-\infty < t < \infty} \max_{1 \leq l \leq d} E|\varepsilon_t(l)|^{2\gamma} < \infty,
\]

for some \( \gamma > 1 \). Equation (3.9) plays an indispensable role in dealing with \( E(y_{n+1} - \hat{y}_{n+1})^2 \) because it is not possible to rigorously analyze

\[
n(E(y_{n+1} - \hat{y}_{n+1})^2 - \sigma_{11}) = E(G_{n+1}^{(n)} \hat{R}_n^{-1} n^{-1/2} \sum_{t=1}^n G_t^{(n)} \varepsilon_t)^2,
\]

without recourse to the moment bounds associated with \( \hat{R}_n^{-1} \) or \( \lambda_{\min}^{-1}(\hat{R}_n) \). Recall \( \sigma_{11} = E(\varepsilon_t^2) \) defined after (2.7). The main result of this paper is stated in the next theorem.

**Theorem 3.1** Assume (3.1), (3.3), (3.4), (2.10), with \( m = d + 1 \), and (3.10), with \( \gamma > 4 \).
Also assume that there exists a $p \times p$ non-random matrix $Q_n$ such that (2.13) and (2.14) hold with $x_t^{(n)} = Q_n s_t$. Then,

\[
\begin{align*}
&n \left[ \mathbb{E}(y_{n+1} - \hat{y}_{n+1})^2 - \sigma_{11} \right] + o(1) = x_{n+1}^{(n)} (n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top})^{-1} x_{n+1}^{(n)} \sigma_{11} \\
&\quad + \sigma_{11} \sum_{j=0}^{d} k_{j},
\end{align*}
\]

(3.12)

Ignoring the $o(1)$ term, the centered MSPE, $\mathbb{E}(y_{n+1} - \hat{y}_{n+1})^2 - \sigma_{11}$, multiplied by the sample size, can be expressed as the sum of two terms. The second term, accounting for the variation due to estimating the ARX part of the model, is linearly proportional to the number of the estimated parameters. On the other hand, the first term (due to the error arising from estimating the deterministic trend) has asymptotic behavior varying appreciably, depending on the time trend’s feature. The following examples provide more illustrations of Theorem 3.1.

**Example 3.1** Consider the polynomial trend (3.6). Set $Q_n = \text{Diag}(1, n^{-1}, \ldots, n^{-p+1})$. Then, $n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top} \rightarrow H_p = (1/(i+j-1))_{1 \leq i, j \leq p}$, the $p$-dimensional Hilbert matrix, and $x_{n+1}^{(n)} \rightarrow 1_p$, the $p$-dimensional vector of ones. Since $H_p^{-1}$ exists (see Choi (1983)), Theorem 3.1 implies

\[
\lim_{n \to \infty} n \left[ \mathbb{E}(y_{n+1} - \hat{y}_{n+1})^2 - \sigma_{11} \right] = \sigma_{11} (1_p^\top H_p^{-1} 1_p + \sum_{j=0}^{d} k_{j}).
\]

(3.13)
Example 3.2  For the periodic trend (3.7), set $Q_n = I_{2h+1}$. Then,

$$
\lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top} = \text{Diag}(1, 1/2, \ldots, 1/2),
$$

and $x_{n+1}^{(n)} = (1, \sin \nu_1(n + 1), \cos \nu_1(n + 1) \ldots, \sin \nu_h(n + 1), \cos \nu_h(n + 1))^\top$. Therefore, $x_{n+1}^{(n)\top} (n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top})^{-1} x_{n+1}^{(n)} \to 2h + 1$, and hence by Theorem 3.1,

$$
\lim_{n \to \infty} n \left[ \mathbb{E}(y_{n+1} - \hat{y}_{n+1})^2 - \sigma_{11} \right] = \sigma_{11}(2h + 1 + \sum_{j=0}^{d} k_j). \quad (3.14)
$$

Example 3.2 reveals that the impact of the periodic trend aligns with that of the ARX component. That is, it is linearly proportional to the number of parameters. On the other hand, Example 3.1 shows that this is not the case for the polynomial trend since $1^\top H_p^{-1} 1_p/p$ is not a constant. We will explore this issue further in the next section.

When $d = 0$ (no exogenous variables in the model), (3.13) was given in Ing (2003) under the stringent condition that $\mathbb{E} |\varepsilon_t|^q < \infty$ for any $q > 0$. His derivation depends on Lemma B.1, which claims that (3.9) holds for any $q > 0$, provided the time trend satisfies (3.6). However, a rigorous proof of this result seems to be missing.

Before closing this section, we note that the relevance of our asymptotic results in finite samples has been investigated through a limited simulation study, which concentrates on several AR(2) models with polynomial or periodic trends. Our simulations show that the empirical MSPEs, obtained based on 10,000 replications, are quite close to their limiting values given on the right-hand sides of (3.13) and (3.14) even for $n = 100$. Further details are available upon request from the authors.
3.2 Statistical predictions and Cauchy matrices

To get a better understanding of the impact of the polynomial time trend on the corresponding MSPE, it is of interest to calculate the value of $1_p^\top H_p^{-1}1_p$ in (3.13). In fact, $H_p$ is a special case of the symmetric Cauchy matrix $C_p = ((l_i + l_j - 1)^{-1})_{1 \leq i, j \leq p}$, where $l_1 \ldots l_p$ are distinct real numbers with $l_i + l_j \neq 1$ for all $1 \leq i, j \leq p$. In this section, we assume $\min_{1 \leq i \leq p} l_i > 1/2$ to ensure that $C_p$ is positive definite; see Fiedler (2010). Obviously, when $l_i = i, i = 1, \ldots, p, C_p = H_p$. By making use of Lagrange’s interpolation formula, Schechter (1959) showed that

$$1_p^\top C_p^{-1}1_p = \left( \sum_{j=1}^{p} 2l_j \right) - p,$$

(3.15)

which can be applied to aerodynamics. Equation (3.15) leads immediately to

$$1_p^\top H_p^{-1}1_p = p^2,$$

(3.16)

showing that estimating the polynomial trend yields a prediction error quadratically proportional to the number of parameters associated with the trend. This is in contrast to the prediction error contributed by estimating the ARX part, which is linearly proportional to the number of parameters. In view of the connection between $1_p^\top H_p^{-1}1_p$ and statistical prediction, it is intriguing to explore if there exists a statistical proof of (3.16) or even (3.15). In the rest of this section, we show that the answer to this question is affirmative. Our proof of (3.15) ((3.16)) is reliant on a novel link between the MSPE and APE.
To begin with, let us focus on the following regression model,

$$ y_t = \sum_{j=1}^{p} \beta_j t_i^{j-1} + \varepsilon_t, \ t = 1, \ldots, n $$

(3.17)

where $l_i > 1/2, i = 1, \ldots, p$, and $\varepsilon_t$ are independent standard normal random variables. The least squares predictor, $\hat{y}_{n+1}$, of $y_{n+1}$ is given by $x_{n+1}^\top \hat{\beta}_n$, where

$$ \hat{\beta}_t = \left( \sum_{j=1}^{t} x_j x_j^\top \right)^{-1} \sum_{j=1}^{t} x_j y_j, $$

with $x_t = (t_i^{t-1}, \ldots, t_i^{p-1})^\top$. Define $D_n = \text{Diag}(n_l^{t-1/2}, \ldots, n_l^{p-1/2})$. Then, by the positive definiteness of $C_p$, it can be shown that

$$ \lim_{n \to \infty} n \{ E(y_{n+1} - \hat{y}_{n+1})^2 - 1 \} = \lim_{n \to \infty} n \{ E(y_{n+1} - \hat{y}_{n+1} - \varepsilon_{n+1})^2 \} $$

$$ = \lim_{n \to \infty} n x_{n+1}^\top D_n^{-1} \left( D_n^{-1} \sum_{t=1}^{n} x_t x_t^\top D_n^{-1} \right)^{-1} D_n^{-1} x_{n+1} = 1_p C_p^{-1} 1_p, $$

(3.18)

which establishes a connection between the left-hand side of (3.15) and the MSPE. The key idea is to further associate the MSPE in (3.18) with the APE, $\sum_{t=M+1}^{n} (y_t - \hat{y}_t - \varepsilon_t)^2$. More specifically, it follows from

$$ x_n^\top \left( \sum_{t=1}^{n} x_n x_n^\top \right)^{-1} x_n \to 0, \ \liminf_{n \to \infty} \lambda_{\text{min}}(D_n^{-1} \sum_{t=1}^{n} x_n x_n^\top D_n^{-1}) > 0, $$

17
Theorem 3 of Wei (1987), and the positive definiteness of $C_p$ that
\[
\lim_{n \to \infty} \frac{\sum_{t=M+1}^{n}(y_t - \hat{y}_t - \varepsilon_t)^2}{\log \det \sum_{t=1}^{n} x_n x_n^\top} = \lim_{n \to \infty} \frac{\sum_{t=M+1}^{n}(y_t - \hat{y}_t - \varepsilon_t)^2}{[(\sum_{j=1}^{p} 2l_j) - p] \log n} = 1 \text{ a.s.,} \tag{3.19}
\]
where $M$ is the smallest integer $t$ such that $\hat{\beta}_t$ is uniquely defined. By Minkowski’s inequality, it can be shown that $\{\sum_{t=M+1}^{n}(y_t - \hat{y}_t - \varepsilon_t)^2/\log n\}$ is uniformly integrable, which, together with (3.19), implies
\[
\lim_{n \to \infty} \frac{1}{\log n} \sum_{t=M+1}^{n} E((y_t - \hat{y}_t - \varepsilon_t)^2) = (\sum_{j=1}^{p} 2l_j) - p. \tag{3.20}
\]
Denote $E((y_t - \hat{y}_t - \varepsilon_t)^2)$ by $\nu_t$. Then, (3.18) and (3.20) ensure
\[
\lim_{n \to \infty} n\nu_n = 1_p^\top C_p^{-1} 1_p \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{\log n} \sum_{t=M+1}^{n} \nu_t = (\sum_{j=1}^{p} 2l_j) - p, \tag{3.21}
\]
respectively. Moreover, it follows from the first identity of (3.21) that
\[
\frac{1}{\log n} \sum_{t=M+1}^{n} \nu_t = \frac{1}{\log n} \left\{ \sum_{t=M+1}^{n} t^{-1}(tv_t - 1_p^\top C_p^{-1} 1_p) + 1_p^\top C_p^{-1} 1_p \sum_{t=M+1}^{n} t^{-1} \right\} = 1_p^\top C_p^{-1} 1_p + o(1),
\]
which, together with the second identity of (3.21), yields (3.15).
4 Conclusion

In this paper, we provide a rigorous analysis of the MSPE of the least squares predictor in ARX models with deterministic time trends satisfying some general conditions. Due to the difficulty in proving moment bounds for $\lambda_{\min}^{-q}(\hat{R}_n), q \geq 1$, the asymptotic expression, (3.12), has not been reported elsewhere to the best of our knowledge. In the polynomial time trend, (3.12) inspires an intriguing proof of the formula for $1_p^T C_p^{-1} 1_p$ from a prediction perspective. However, there are still several issues that require further investigation. For example, both the polynomial and periodic time trends, (3.6) and (3.7), are precluded by (3.4) if $1 - a_1 z - \cdots - a_k z^k$ has roots on the unit circle. This leads to the question on how to modify (3.13) and (3.14) in the presence of unit-roots. The techniques developed in Chan (1989) may be helpful to answer this question. Also, since the models imposed on the exogenous variables $z_t(l)s$ are very general, the multistep prediction based on a finite number of lags of $z_t(l)$ may result in model misspecification. Therefore, an extension of (3.12) to the case of multistep prediction or model misspecification is another interesting topic for future research.

A Appendix

A.1 Proof of Theorem 2.1

To prove Theorem 2.1, we need some technical lemmas to characterize (2.14).

Lemma A.1 Assume (2.13). Then, (2.14) holds if and only if there exists a subset, $G_n$,
of $X_n = \{1, \ldots, n\}$, with $\liminf_{n \to \infty} \sharp(G_n)/n > 0$ and $\liminf_{n \to \infty} \min_{t \in G_n} \|x_t^{(n)}\| > 0$ such that

$$\liminf_{n \to \infty} \lambda_{\min}(\sharp(G_n)^{-1} \sum_{t \in G_n} x_t^{(n)} x_t^{(n)\top}) > 0.$$  \hfill (A.1)

**Proof.** The proof of the "if" part of Lemma A.1 is easy and hence omitted. To prove the "only if" part, we note that (2.14) yields that for all large $n$,

$$\lambda_{\min}(n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top}) > s,$$  \hfill (A.2)

where $s$ is some positive constant, and hence

$$n^{-1} \sum_{t=1}^{n} x_{t,1}^2 > s,$$  \hfill (A.3)

where $x_{t,1}$ denotes the ith component of $x_t^{(n)}$. Therefore, $G_n := \{t : 1 \leq t \leq n, \|x_t^{(n)}\| > s/2\}$ is non-empty for all large $n$. By (2.13) and (A.3), one has for all large $n$, $ns \leq \sum_{t=1}^{n} x_{t,1}^2 \leq \sharp(G_n) M_1^2 + ns/2$, yielding

$$\sharp(G_n) \geq sn/(2M_1^2).$$  \hfill (A.4)

Now, the desired conclusion follows from (A.4), (A.2), and $\lambda_{\min}(\sharp(G_n)^{-1} \sum_{t \in G_n} x_t^{(n)} x_t^{(n)\top}) \geq \lambda_{\min}(n^{-1} \sum_{t \in G_n} x_t^{(n)} x_t^{(n)\top}) \geq \lambda_{\min}(n^{-1} \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top}) - n^{-1} \sum_{t=1, t \notin G_n} \|x_t^{(n)}\|^2$.

**Lemma A.2** Assume (2.13). Then, (2.14) holds if and only if there exist disjoint subsets,
$D_1, \ldots D_{q_n}$, of $X_n$, where $|D_i| = p$, $i = 1, \ldots, q_n$, and $\liminf_{n \to \infty} q_n/n > 0$, such that

$$\liminf_{n \to \infty} \min_{1 \leq i \leq q_n} \lambda_{\min} \left( \sum_{t \in D_i} x_i^{(n)} x_i^{(n)\top} \right) > 0.$$  \hfill (A.5)

**Proof.** The proof of the "if" part of Lemma A.2 is easy and hence omitted. To prove the "only if" part, by Lemma A.1 and (2.13), we can set $\|x_t^{(n)}\| = 1$ for all $t$. We also assume without loss of generality that the $s$ in (A.2) is less than 1. Define $D_0(n) = \emptyset$, and for $i \geq 1$, let $D_i(n)$ be any $p$-element subset of $X_n - \bigcup_{l=0}^{i-1} D_l(n)$ satisfying

$$\lambda_{\min} \left( \sum_{t \in D_i(n)} x_t^{(n)} x_t^{(n)\top} \right) > \frac{s^p}{2p^{p-1}},$$  \hfill (A.6)

and $D_i(n) = \emptyset$ if no such subset is found. Also define $q_n = 0$ if $D_1(n) = \emptyset$, and $\max\{i \geq 1 : D_i(n) \neq \emptyset\}$ otherwise. For notational simplicity, in what follows we suppress the dependence of $x_t^{(n)}$ and $D_i(n)$ on $n$, and write $x_t$ and $D_i$ instead of $x_t^{(n)}$ and $D_i(n)$, respectively. Denote $X_n - \bigcup_{l=0}^{q_n} D_l$ by $Z_n = \{s_1, \ldots, s_{k_n}\}$, where $k_n = n - pq_n$. If $k_n < p$, then

$$q_n > n/p - 1.$$  \hfill (A.7)

For $k_n \geq p$, choose distinct elements, $c_1, \ldots, c_p$, in $Z_n$ sequentially as follows. Let $c_1$ be any element of $Z_n$. For $2 \leq j \leq p$, set $c_j = \arg\max_{s_i \in Z_n} \|(I_p - M_{j-1})x_{s_i}\|$, where $M_{j-1}$ is the orthogonal projection matrix onto $C(x_{c_1}, \ldots, x_{c_{j-1}})$, the column space of $(x_{c_1}, \ldots, x_{c_{j-1}})$. Note that this sequential procedure of choosing $c_i$ is similar in spirit to
the orthogonal greedy algorithm in Ing and Lai (2011). Let $M_0$ be the $p \times p$ matrix of zeros. Then,

$$\| (I_p - M_{j-1}) \mathbf{x}_{c_j} \| \text{ is non-increasing in } j, \quad (A.8)$$

and

$$\prod_{j=1}^{p} \| (I_p - M_{j-1}) \mathbf{x}_{c_j} \|^2 \leq \lambda_{\min}(\sum_{j=1}^{p} \mathbf{x}_{c_j} \mathbf{x}_{c_j}^\top) \lambda_{\max}^{p-1}(\sum_{j=1}^{p} \mathbf{x}_{c_j} \mathbf{x}_{c_j}^\top) \leq \lambda_{\min}(\sum_{j=1}^{p} \mathbf{x}_{c_j} \mathbf{x}_{c_j}^\top)p^{p-1} \leq (s/2)^p, \quad (A.9)$$

where $\lambda_{\max}(A)$ denotes the maximum eigenvalue of $A$ and the last inequality is ensured by

$$\max_{D \subset \mathbb{Z}_n : \#(D) = p} \lambda_{\min}(\sum_{t \in D} \mathbf{x}_t \mathbf{x}_t^\top) \leq \frac{s^p}{2^{pp-1}}.$$ 

Equations (A.8) and (A.9) imply there exists $1 < j^* \leq p$ such that for all $j^* \leq j \leq p$ and all large $n$, $\| (I_p - M_{j-1}) \mathbf{x}_{c_j} \|^2 \leq s/2$, and hence for all $1 \leq i \leq k_n$ and all large $n$,

$$\| (I_p - M_{j^*-1}) \mathbf{x}_{s_i} \|^2 \leq s/2. \quad (A.10)$$
Let \( \mathbf{v} \) be any unit vector perpendicular to \( C(M_{j^*-1}) \). Then, (A.10) yields

\[
\mathbf{v}^\top (n^{-1} \sum_{i=1}^{k_n} \mathbf{x}_i \mathbf{x}_i^\top) \mathbf{v} \leq n^{-1} \sum_{i=1}^{k_n} [\mathbf{x}_i^\top (I_p - M_{j^*-1}) \mathbf{v}]^2 \leq s/2,
\]

and hence \( \lambda_{\min}(n^{-1} \sum_{i=1}^{k_n} \mathbf{x}_i \mathbf{x}_i^\top) \leq s/2 \). This, together with (A.2) and

\[
\lambda_{\min}(n^{-1} \sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i^\top) = \lambda_{\min}(n^{-1} \sum_{i=0}^{q_n} \sum_{l \in D_i} \mathbf{x}_i^\top \mathbf{x}_l^\top + n^{-1} \sum_{j=1}^{k_n} \mathbf{x}_j \mathbf{x}_j^\top)
\]

\[
\leq \frac{q_n p}{n} + \lambda_{\min}(n^{-1} \sum_{j=1}^{k_n} \mathbf{x}_j \mathbf{x}_j^\top),
\]

gives

\[
q_n \geq \frac{sn}{2p},
\]

(A.11)

for all large \( n \). Consequently, the desired conclusion follows from (A.7) and (A.11).

With the help of Lemma A.2, we are now in a position to prove Theorem 2.1.

**Proof of Theorem 2.1.** Consider

\[
\omega_t^* = \begin{pmatrix} I_p & 0' \\ 0 & C_0^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{x}_t^{(n)} \\ \mathbf{Y}_t \end{pmatrix} \equiv \begin{pmatrix} \mathbf{x}_t^{(n)} \\ \mathbf{Y}_t^* \end{pmatrix},
\]

where \( \mathbf{Y}_t^* = \mathbf{e}_{t,0} + \sum_{j=1}^{\infty} C_j^* \mathbf{e}_{t,j} \), \( C_j = C_0 C_j \), and the dependence of \( \omega_t^* \) on \( n \) is suppressed.
in this proof. It follows from (2.2) that

$$\sum_{j=1}^{\infty} \|C_j^*\|_{F}^2 < \infty.$$  \hfill (A.12)

In the rest of the proof, we shall show that

$$E \left( \lambda_{\min}^{-q} \left( n^{-1} \sum_{t=1}^{n} \omega_t \omega_t^\top \right) \right) = O(1),$$  \hfill (A.13)

which leads immediately to the desired result (2.12). By Lemma A.2, one has for all large $n$,

$$\min_{1 \leq i \leq q_n} \lambda_{\min} \left( \sum_{t \in D_i} x_t^{(n)} x_t^{(n)\top} \right) > \rho_1,$$  \hfill (A.14)

where $\rho_1$ is some positive constant, $D_i$s are disjoint subsets of $X_n$ with $\|D_i\| = p$, for all $1 \leq i \leq q_n$, and $q_n$ satisfies $\lim \inf_{n \to \infty} q_n/n > 0$. Let $d_i$ denote the largest integer in $D_i$ and $\{d_{(i)}\}$ be the decreasing rearrangement of $\{d_i\}$. Set $c_1 = d_{(1)}$ and for $i \geq 2$, define $c_i = \max\{d_{(l)}, 1 \leq l \leq q_n : c_{i-1} - d_{(l)} \geq k\}$, and 0 if no such $d_{(l)}$ exists. Let $s_n = \max\{i \geq 1 : c_i \geq 1\}$. Then, it is easy to see that $\lim \inf_{n \to \infty} s_n/n > 0$. Let $D'_i$ denote the set $D_j, 1 \leq j \leq q_n$, containing $c_i$, $L$ be an integer satisfying

$$L > \frac{2 + (q^{-1} + \gamma^{s-1})\iota}{\alpha(q^{-1} - \gamma^{s-1})},$$  \hfill (A.15)

and $g_n = \lfloor s_n/L \rfloor$, where $q < \gamma < \gamma^*$, $\iota = p + km$, $\alpha$ is defined in (2.10), and $\lfloor a \rfloor$ is the largest integer $\leq a$. Then, by the convexity of $x^{-q}, x > 0$, and $\lim \inf_{n \to \infty} g_n/n > 0$, we
obtain for all large \( n \),

\[
\text{the left-hand side of (A.13)} \leq \mathbb{E}\left( \lambda_{\min}^{-q} \left( n^{-1} \sum_{i=1}^{g_n} \sum_{t \in D'_i} B_t \right) \right)
\]

\[
\leq C g_n^{-1} \sum_{j=0}^{g_n-1} \mathbb{E}\left( \lambda_{\min}^{-q} \left( \sum_{i=1}^{L} \sum_{t \in D'_{i+jL}} B_t \right) \right),
\]

(A.16)

where \( B_t = \omega_t^* \omega_t^{\top} \) and \( C \) here and hereafter represents a generic positive constant which is independent of \( n \), but may vary from place to place. In view of (A.16), (A.13) follows if we can show that for all large \( n \),

\[
\mathbb{E}\left( \lambda_{\min}^{-q} \sum_{i=1}^{L} \sum_{t \in D'_{i+jL}} B_t \right) \leq C, \quad j = 0, \ldots, g_n - 1.
\]

(A.17)

In the following, we only prove (A.17) for the case of \( j = 0 \) because the proofs of other cases can be obtained similarly.

Let \( k^* = \max\{k_1^*, k_2^*\} \), where \( k_1^* \) and \( k_2^* \) are positive constants to be specified later. Then, the left-hand sides of (A.17) (with \( j = 0 \)) is bounded by

\[
k^* + \int_{k^*}^{\infty} P(G(\mu) \cap V(\mu)) d\mu + \int_{k^*}^{\infty} P(V^c(\mu)) d\mu
\]

\[
:= k^* + \int_{k^*}^{\infty} I(\mu) d\mu + \int_{k^*}^{\infty} II(\mu) d\mu,
\]

(A.18)

where

\[
G(\mu) = \left\{ \inf_{\|\nu\|=1} \sum_{i=1}^{L} \sum_{t \in D'_i} (\nu' \omega_t^*)^2 < \mu^{-1/q} \right\}.
\]
and
\[
V(\mu) = \left\{ \max_{t \in \bigcup_{i=1}^{L_i} D'_i} \|\omega_i^*\|^2 \leq s\mu^{1/\gamma^*} \right\},
\]
in which \(s\) is small enough such that
\[
2M_1 p\sqrt{s} + ps \leq \rho_1/4, \tag{A.19}
\]
where \(M_1\) is defined in (2.13). By (2.15), (A.12), and Lemma 2 of Wei (1987), it can be shown that
\[
\int_{k^*}^\infty \Pi(\mu)d\mu \leq C \int_{k^*}^\infty \mu^{-\gamma/\gamma^*} d\mu = O(1). \tag{A.20}
\]

To deal with the first integration in (A.18), note that
\[
G(\mu) \cap V(\mu) \subset \bigcap_{i=1}^{L_i} \left\{ \inf_{\nu \in \mathbb{R}^i} \sum_{t \in D'_i} (\nu'\omega_i^*)^2 < \mu^{-1/q}, \sum_{t \in D'_i} \|\omega_i^*\|^2 \leq ps\mu^{1/\gamma^*} \right\} \cap V(\mu)
\]
\[:= Q(\mu) \cap V(\mu).\]

By an argument similar to that used in Ing and Wei (2003, page 137), it can be shown that there exist a positive integer \(m^* \leq C_1\mu^{(2^{-1} + \gamma^*-1)/2}\) and unit vectors, \(l_1, \ldots, l_{m^*}\), in \(\mathbb{R}^n\) such that
\[
Q(\mu) \subset \bigcup_{j=1}^{m^*} \{\|l_j\sum_{t \in D'_i} B_t \leq (2\sqrt{ps} + 1)\mu^{-1/2q}, i = 1, \ldots, L\} := \bigcup_{j=1}^{m^*} \Pi_{j,L}(\mu),
\]
where \(C_1\) is some positive constant independent of \(n\) and \(\mu\) and \(\|x\|_A^2 = x^TAx\) for non-
negative definite matrix $A$. Since $|U_j^T \omega^*_c| \leq \|U_j\|_{\sum_{t \in D_t} B_t}$,

\[ I(\mu) \leq \sum_{j=1}^{m^*} P \left( |U_j^T \omega^*_c| \leq \left(2\sqrt{psl} + 1\right)\mu^{-1/2q}, \ i = 1, \ldots, L \right) \]

\[ + \sum_{j=1}^{m^*} P \left( V(\mu), \Pi_{j,L}(\mu) \right) := \sum_{j=1}^{m^*} III_j(\mu) \]

\[ + \sum_{j=1}^{m^*} IV_j(\mu), \]

where $l_{j,2}$ is the vector consisting of $l_j$'s last $km$ elements. Denote $l_{j,2}$ by $(U_{j,2}(1), \ldots, U_{j,2}(k))^T$, where each of $l_{j,2}(i)$ is $m$-dimensional. Then, for $\|l_{j,2}\| \geq \mu^{-1/(2\gamma^*)}$ and $\mu \geq k^*_1 = \{2\sqrt{k}(2\sqrt{psl} + 1)/\delta\}^{2/(q^{-1} - \gamma^{-1})}$, it holds that

\[ III_j(\mu) = E \left\{ \left. \prod_{i=2}^L I_{|l_j^T \omega^*_c| \leq (2\sqrt{psl} + 1)\mu^{-1/2q}} \right| \mathbf{A}_1(\mu) \right\} \]

\[ \leq M \{2\sqrt{k}(2\sqrt{psl} + 1)\mu^{-(q^{-1} - \gamma^{-1})/2}\} \sigma E \left( \prod_{i=2}^L I_{|l_j^T \omega^*_c| \leq (2\sqrt{psl} + 1)\mu^{-1/2q}} \right) \]

where $\mathbf{A}_1(\mu) = \{-(2\sqrt{psl} + 1)\mu^{-1/2q} - r_1^* \leq \sum_{h=1}^k l_{j,2}(h) \delta_{c_{1-h}C_{c_{1-h}}} \leq (2\sqrt{psl} + 1)\mu^{-1/2q} - r_1^* \}$, with $r_1^* = l_{j,1}^l x_{c_1}^{(n)} + l_{j,2} l_{j,2}^T \sum_{j=1}^{\infty} C_{c_1,j}^* \varepsilon_{c_1,j}$ and $(l_{j,1}^T, l_{j,2}^T)^T = l_j$, and the first inequality follows from $\|l_{j,2}(h)\| \geq k^{-1/2}\mu^{-1/(2\gamma^*)}$ for some $1 \leq h \leq k$, (2.10), and the independence among
\{δ_t\}. Repeat the same argument \(L - 1\) times, one gets,

\[
\text{III}_j(\mu) \leq M^L \{2\sqrt{k(2\sqrt{psl} + 1)}\}^{\nu L} \mu^{-(q^{-1} - \gamma^{-1})\alpha L/2},
\]

provided \(\|l_{j,2}\| \geq \mu^{-1/(2\gamma^*)}\) and \(\mu \geq k_1^*\). As a result, by (A.15),

\[
\int_{k^*}^{\infty} \sum_{j=1}^{m^*} \text{III}_j(\mu) d\mu \leq C \int_{k^*}^{\infty} \mu^{-(q^{-1} - \gamma^{-1})\alpha L/2 - (q^{-1} + \gamma^{-1})/2} = O(1). \tag{A.22}
\]

For \(\|l_{j,2}\| < \mu^{-1/(2\gamma^*)}\) and \(\mu \geq k_2^* = \text{max\{2\gamma^*, \{5(2\sqrt{psl} + 1)/\rho_1\}^q\}}\), (A.14) and (A.19) ensure that on the set \(V(\mu)\),

\[
\min_{1 \leq i \leq L} \|l_i\| \sum_{t \in D_i} B_t \geq \left( \min_{1 \leq i \leq L} \|l_{i,1}\|^2 \sum_{t \in D_i} x_i^{(n)} x_i^{(n)\top} - 2M_1 p \sqrt{s - ps}\right)^{1/2} \\
\geq (\rho_1/2 - 2M_1 p \sqrt{s - ps})^{1/2} \geq (\rho_1/4)^{1/2} \geq (\rho_1/5)^{1/2} \geq (2\sqrt{psl} + 1)\mu^{-1/2q},
\]

for all large \(n\). Hence, for all large \(n\),

\[
\int_{k^*}^{\infty} \sum_{j=1}^{m^*} \text{IV}_j(\mu) d\mu = 0. \tag{A.23}
\]

Consequently, (A.17) (with \(j = 0\)) follows from (A.18), (A.20), (A.21), (A.22), and (A.23). Thus, the proof is complete.
A.2 Proof of Theorem 3.1

PROOF OF THEOREM 3.1. We can assume without loss of generality that \( \hat{R}_n^{-1} \) exists because (3.9) implies \( P(\hat{R}_n^{-1} \text{ exists}) = 1 \) for all large \( n \). Denote \( (v^\top_t(k_0), z^\top_t(k_1), \ldots, z^\top_t(k_d))^\top \) by \( Q_t = (Q_t(1), \ldots, Q_t(\sum_{l=0}^d k_l))^\top \) and \( \mathbb{E}(Q_tQ_t^\top) \) by \( F = (F_{i,j})_{1 \leq i,j \leq \sum_{l=0}^d k_l} \). Then, it follows from (2.6), (2.7), and the first moment bound theorem of Findley and Wei (1993) that for any \( 1 \leq i,j \leq \sum_{l=0}^d k_l \),

\[
\mathbb{E}(n^{-1} \sum_{t=1}^n Q_t(i)Q_t(j) - F_{i,j})^2 \leq C \left( \sum_{l=0}^{n-1} \gamma_l^2(i)/n, \sum_{l=0}^{n-1} \gamma_l^2(j)/n \right) = o(1), \tag{A.24}
\]

where \( \gamma_l(i) \) is the autocovariance function of \( \{Q_t(i)\} \) at lag \( l \). In addition, it is easy to see that for any \( 1 \leq i \leq \sum_{l=0}^d k_l \),

\[
\mathbb{E}\|n^{-1} \sum_{t=1}^n x_t(n)^\top Q_t(i)\|^2 \to 0,
\]

which, together with (A.24), yields

\[
\hat{R}_n = \begin{pmatrix} n^{-1} \sum_{t=1}^{n-1} x_t(n)^\top x_t(n) & 0_{p \times \sum_{l=0}^d k_l} \\ 0_{(\sum_{l=0}^d k_l) \times p} & F \end{pmatrix} = o_p(1). \tag{A.25}
\]

In view of (A.25), the desired conclusion follows if

\[
T_n := (G_{n+1}^{(n)} \hat{R}_n^{-1} n^{-1/2} \sum_{t=1}^n G_t^{(n)} \xi_t)^2 \text{ is uniformly integrable,} \tag{A.26}
\]
and
\[
\mathbb{E}\left(\sum_{t=1}^{n-1} x_t^{(n)} \left( x_t^{(n)} \right)^{\top} - n \frac{1}{2} \sum_{t=1}^{n} x_t^{(n)} \varepsilon_t + Q_n^{\top} F_n^{-1} \sum_{t=1}^{n} Q_t \varepsilon_t\right)^2
\]
\[
= x_{n+1}^{(n)} \left( \sum_{t=1}^{n} x_t^{(n)} \left( x_t^{(n)} \right)^{\top} - n \frac{1}{2} \sum_{t=1}^{n} x_t^{(n)} \varepsilon_t \right) \sigma_{11} + \sigma_{11} \sum_{j=0}^{d} k_j + o(1).
\]

(A.27)

Since $\gamma > 4$, there exists $\theta > 0$ small enough such that $4 < 2 \gamma (1 + \theta) / (\gamma - 2(1 + \theta))$. Let $2 \gamma (1 + \theta) / (\gamma - 2(1 + \theta)) \leq q^* < \gamma$. Then, by Theorem 2.1 and (3.10),

\[
\mathbb{E}\{\lambda_{\min}^{-q^*} (\mathbf{R}_n)\} = O(1).
\]

(A.28)

By (A.28), Lemma 2 of Wei (1987), $4 q^* (1 + \theta) / (q^* - 2 (1 + \theta)) \leq 2 \gamma$, and Hölder’s inequality,

\[
\mathbb{E}(T_n^{1+\theta}) \leq \left( \mathbb{E}\|G_{n+1}^{(n)}\|^{4 q^* (1 + \theta) / 2 q^*} \right)^{\frac{q^* - 2 (1 + \theta)}{q^* - 2 (1 + \theta)}} \left( \mathbb{E}\|G_{n+1}^{(n)}\|^{4 q^* (1 + \theta) / 2 q^*} \right)^{\frac{2 (1 + \theta)}{2 q^*}}
\]

leading to (A.26).

To prove (A.27), define $Q_n = \mathbb{E}(Q_n | \mathcal{F}_{n-g_n})$, where $g_n \to \infty$, $g_n = o(n)$, and $\mathcal{F}_t$ is the
σ-field generated by \((\delta_t, \delta_{t-1}, \ldots)\). Then,

\[
\begin{align*}
\mathbb{E} \left( Q_n^{\top} F^{-1} n^{-1/2} \sum_{t=1}^{n} Q_{t-1} \varepsilon_t \right)^2 &= \mathbb{E} \left( Q_n^{\star \top} F^{-1} n^{-1/2} \sum_{t=1}^{n-g_n} Q_{t-1} \varepsilon_t \right)^2 \\
+ \mathbb{E} \left( \tilde{Q}_n^{\top} F^{-1} n^{-1/2} \sum_{t=n-g_n+1}^{n} Q_{t-1} \varepsilon_t \right)^2 + \mathbb{E} \left( Q_n^{\star \top} F^{-1} n^{-1/2} \sum_{t=n-g_n+1}^{n} Q_{t-1} \varepsilon_t \right)^2 \\
+ \mathbb{E} \left( \tilde{Q}_n^{\top} F^{-1} n^{-1/2} \sum_{t=1}^{n-g_n} Q_{t-1} \varepsilon_t \right)^2 := (I) + (II) + (III) + (IV),
\end{align*}
\]

where \(Q_n^{\star} = Q_n - \tilde{Q}_n\). By Lemma 2 of Wei (1987) and the Cauchy-Schwarz inequality,

\[
(II) + (III) + (IV) = o(1).
\]

Straightforward calculations give

\[
\lim_{n \to \infty} (I) = \sigma_{11} \sum_{j=0}^{d} k_j,
\]

and

\[
\begin{align*}
\mathbb{E} \left( x_t^{(n)\top} (n-1 \sum_{t=1}^{n-1} x_t^{(n)} x_t^{(n)\top})^{-1} n^{-1/2} \sum_{t=1}^{n} x_t^{(n)} \varepsilon_t \right)^2 \\
= x_{n+1}^{(n)\top} (n-1 \sum_{t=1}^{n} x_t^{(n)} x_t^{(n)\top})^{-1} x_{n+1}^{(n)} \sigma_{11}
\end{align*}
\]
An argument similar to that used to prove (A.30) yields

\[
E \left( n^{-1} \{ x_t^{(n)^\top} (n^{-1} \sum_{t=1}^{n-1} x_t^{(n)} x_t^{(n)^\top})^{-1} \sum_{t=1}^{n} x_t^{(n)} \epsilon_t \} (Q_n^\top F^{-1} \sum_{t=1}^{n} Q_{t-1} \epsilon_t) \right) = o(1). \quad (A.33)
\]

Consequently, (A.27) follows from (A.29)–(A.33). Thus the proof is complete.

**Acknowledgement**

We would like to thank the Associate Editor and two anonymous reviewers for their comments and suggestions that helped improve the readability of this paper. Ing’s and Yu’s research was supported in part by grants 109-2118-M-007-007-MY3 and 109-2118-M-390-002 from the Ministry of Science and Technology (MOST), Taiwan.

**References**


