Resampling-based Bias-corrected Time Series Prediction

S. Bandyopadhyay and S. N. Lahiri

Department of Statistics

Texas A & M University

College Station, TX 77843-3143

Running Head: Resampling based prediction

ABSTRACT

In this paper, we consider estimation of the mean squared prediction error (MSPE) of the best linear predictor of (possibly) nonlinear functions of finitely many future observations in a stationary time series. We develop a resampling methodology for estimating the MSPE when the unknown parameters in the best linear predictor are estimated. Further, we propose a bias corrected MSPE estimator based on the bootstrap and establish its second order accuracy. Finite sample properties of the method are investigated through a simulation study.

---

*Research partially supported by NSF grant no. DMS 0707139.

AMS (2000) subject classification: Primary 62M30, Secondary 62E20.

Keywords and phrases: Bootstrap, Mean squared prediction error, second order bias correction, tilting.

1 Introduction

Let $\{X_t\}_{t=-\infty}^{\infty}$ be a second order stationary time series with auto-covariance function $\gamma(\cdot)$ and spectral density $f(\cdot)$. Suppose that a finite stretch, $X_1, \ldots, X_n$ of the series is observed. In many applications, it is important to predict an unobserved future value X_{n+k} in the time series or more generally, a suitable functional of a set of future values $X_{n+1}, \ldots, X_{n+k}$:

$$\Psi = \psi(X_{n+1}, \ldots, X_{n+k})$$

where $\psi : I\!\!R^k \to I\!\!R$ is a known function and $k \in I\!\!N$. Here and in the following, $I\!\!N$ and $Z\!\!\!Z$ respectively denote the set of all positive integers and the set of all integers. A popular predictor of Ψ is given by the best linear predictor (BLP) $\tilde{\Psi}_n = \alpha_1 X_1 + \ldots + \alpha_n X_n$, where

$$(\lambda_1, \ldots, \lambda_n) = \text{argmin}_{a_1,\ldots,a_n} E\Big(\Psi - [a_1 X_1 + \ldots + a_n X_n]\Big)^2. \tag{1.1}$$

The co-efficients $\lambda_1, \ldots, \lambda_n$ can be found by standard optimization arguments from calculus; See (2.1), Section 2 below for an explicit expression for $\lambda_1, \ldots, \lambda_n$. Typically, $\lambda_1, \ldots, \lambda_n$ in $\tilde{\Psi}_n$ depend on the auto-covariance function $\gamma(\cdot)$ and, for a nonlinear $\psi(\cdot)$, on other population parameters of the $\{X_t\}$-process and hence, are typically unknown in practice. In this paper, we restrict attention to parametric time series models and highlight the dependence of the BLP on the underlying parameters by writing

$$\tilde{\Psi}_n = \tilde{\Psi}_n(\theta)$$

where $\theta \in I\!\!R^p$ $(p \in I\!\!N)$ is the vector of unknown parameters of the $\{X_t\}$-process. Since $\tilde{\Psi}(\theta)$ depends on unknown θ, it is not usable in practice. A common approach is to plug-in an estimator $\hat{\theta}_n$ of the unknown parameter θ in $\tilde{\Psi}_n(\theta)$, yielding the estimated best linear predictor (EBLP):

$$\hat{\Psi}_n = \tilde{\Psi}_n(\hat{\theta}_n). \tag{1.2}$$

An important problem in time series analysis is to accurately estimate the mean squared prediction error (MSPE) of the EBLP:

$$M(\theta) \equiv E\Big(\hat{\Psi}_n - \Psi_n(\theta)\Big)^2. \tag{1.3}$$

Like the BLP $\tilde{\Psi}_n(\theta)$, the MSPE also depends on the unknown parameter vector θ. Note that the function $M(\theta) \equiv M_n(\theta)$ can be represented as

$$\begin{aligned} M(\theta) &= E\Big(\tilde{\Psi}_n(\theta) - \Psi\Big)^2 + 2E\Big[\{\hat{\Psi}_n - \tilde{\Psi}_n(\theta)\}\{\tilde{\Psi}_n(\theta) - \Psi\}\Big] + E\Big(\hat{\Psi}_n - \tilde{\Psi}_n(\theta)\Big)^2 \\ &\equiv M_1(\theta) + M_2(\theta) + M_3(\theta), \quad \text{say.} \end{aligned} \tag{1.4}$$

The first term $M_1(\theta) \equiv M_{1n}(\theta)$ is the MSPE of the ideal predictor $\tilde{\Psi}_n(\theta)$, the third term $M_3(\theta) \equiv M_{3n}(\theta)$ is the estimation error due to the substitution of $\hat{\theta}_n$ in place of θ in $\tilde{\Psi}_n(\cdot)$, and the second one is a cross-product term. Thus, the MSPE of the EBLP depends on the MSPE of the ideal predictor as well as on the particular estimator $\hat{\theta}_n$ used for estimating the unknown parameter vector θ. Except for some very specific cases, analytic expressions for the functions $M_i(\theta)$, $i = 1, 2, 3$ (particularly, $M_2(\theta)$ and $M_3(\theta)$) are not available in the literature, making the estimation of the MPSE $M(\theta)$ difficult by the traditional plug-in approach. In this paper, we propose a bootstrap based method to derive an estimator of the MPSE $M(\theta)$. The key advantage of the bootstrap methodology is that it produces an estimator of the MSPE of the EBLP *for any given* estimator $\hat{\theta}_n$ of θ, without requiring any analytical computation of the functions $M_2(\theta)$ and $M_3(\theta)$ which critically depend on the choice of $\hat{\theta}_n$. We show that under fairly mild regularity conditions on the $\{X_t\}$-process and on the estimators $\hat{\theta}_n$, the bootstrap MSPE estimator is consistent.

Next we consider higher order accuracy of the resulting bootstrap estimator. Typically, of the three terms $M_i(\theta)$, $i = 1, 2, 3$, the first one is $O(1)$, while the second and the third terms are typically $O(n^{-1})$, as the sample size n goes to infinity. As a result, usual consistency of the "ordinary" bootstrap MSPE estimator of $M(\theta)$ is not adequate in many applications where the sample size only moderately large and the effects of the $O(n^{-1})$ terms can not be ignored. Indeed, it can be shown that the bootstrap MSPE estimator has a bias of the order $O(n^{-1})$, which is of the same order as the orders of the terms $M_i(\theta)$, $i = 2, 3$. Thus, the "ordinary" bootstrap MSPE estimator masks the contributions coming from parameter estimation in $\tilde{\Psi}_n(\theta)$ to the overall MSPE of $\hat{\Psi}_n$. What is needed is an estimator of the MSPE of $\hat{\Psi}_n$ that has a bias of order $o(n^{-1})$ and still retains the standard order of convergence; Following Prasad and Rao (1990), we call such estimators of the MSPE $M(\theta)$ *second order correct*. A common way to construct a second order correct MSPE estimator is to use the explicit bias correction to a plug-in estimator of $M(\theta)$. However, this is impractical and undesirable in our situation mainly because of two reasons, namely, (i) explicit analytical expressions for $M_i(\theta)$, $i = 1, 2, 3$ are very rarely available in the literature (only in some simple toy models) as these are very difficult to derive in reasonable generality, and (ii) the explicit bias correction leads to a negative estimator of the MSPE with a positive probability. An important contribution of the paper is to develop a new method for constructing a second order correct MSPE estimator that is non negative with probability one. The key idea is to "tilt" the estimator $\hat{\theta}_n$ suitably so that it balances out the bias of the "ordinary" bootstrap MSPE estimator to the order $O(n^{-1})$. The tilting factor used here is based on certain iterations of the bootstrap step and on a simple formula to combine them. As a result, the computation of the proposed second order correct MSPE estimator is very much feasible with today's computing power, and the methodology works any choice of the estimator $\hat{\theta}_n$ satisfying the mild regularity conditions of the main result. Most importantly, the proposed method does not require any analytical derivation on the part of the user.

The rest of the paper is organized as follows. We conclude this section with a brief literature review. In Section 2, we describe the "ordinary" bootstrap estimator of the MSPE and prove its consistency. The tilted version of the MSPE estimator and its theoretical properties are stated in Section 3. In Section 4, we develop some bootstrap based approximations for different functions appearing in the tilted MSPE estimator, for which exact analytical expressions are either unavailable or intractable. In Section 5, we report the results from a simulation study on finite sample properties of the proposed tilted MSPE estimator. Proofs of the main results are presented in Section 6.

The literature on time series prediction is huge and is well documented in the case where the target variable $\Psi = X_{n+k}$ for some k; See Brockwell and Davis (1991), Priestley (1981). For standard stationary time series models, like the autoregressive (AR) processes and autoregressive and moving average (ARMA) processes, explicit expressions for $M_1(\theta)$ is known (cf. Brockwell and Davis (1991)), although expressions for $M_2(\theta)$ and $M_3(\theta)$ are not common. The masking-effect of the naive plug-in approach on MSPE estimation was pointed out by Prasad and Rao (1990), who also introduced the concept of second order bias corrected MSPE estimators, in the context of small area estimation. For a detailed account of the literature on issues and solutions in the small area estimation problem until 2003, see Rao (2003). In the time series context, Ansley and Kohn (1986) and Quenneville and Singh (2000) proposed different MSPE estimators based on analytical considerations for the state space model. More recently, Pfeffermann and Tiller (2005) proposed a bootstrap based method for MSPE estimation of the best linear unbiased predictor (BLUP), also for the state-space model under a Gaussian assumption. The second order correct MSPE estimation methodology presented here is different from the earlier work on the problem in the time series literature; It is based on the approach developed by Lahiri and Maiti (2003) and Lahiri et al. (2007) in the context of small area estimation.

2 Bootstrap Estimation of the MSPE

2.1 Preliminaries

In this section, we formalize the basic framework for bootstrap estimation of the MSPE. As in Section 1, let $\{X_t\}_{t=-\infty}^{\infty}$ be a second order stationary time series with an absolutely summable auto-covariance function $\gamma(\cdot)$ and spectral density $f(\cdot)$, and for a known function $\psi : I\!R^k \to I\!R$ let

$$\Psi = \psi(X_{n+1}, \ldots, X_{n+k})$$

is to be predicted using the observations $X_1, \ldots, X_n$. For the ease of exposition and as it is customary in the time series literature (cf. Chapter 5, Brockwell and Davis (1991)), for the rest of this paper, we shall suppose that the variables X_t's and Ψ have mean zero. Thus, the focus of the paper is on the prediction of the random part; The deterministic mean part, if any, can be estimated by any of the standard methods, such as (quasi-)maximum likelihood, method of moments, etc., which in turn, can be used for mean correction. Under the zero-mean assumption, it is easy to derive an explicit expression for the BLP $\tilde{\Psi}_n$ using standard arguments. Thus, by differentiating the expression on the right side of (1.1), it is easy to show that the vector $\boldsymbol{\lambda}_n \equiv (\lambda_1, \ldots, \lambda_n)'$ of co-efficients in $\tilde{\Psi}_n$ are given by

$$\boldsymbol{\lambda}_n \equiv \boldsymbol{\lambda}_n(\theta) = \Gamma_n^{-1}\boldsymbol{\gamma}_n, \tag{2.1}$$

where $\Gamma_n \equiv \Gamma_n(\theta)$ is the $n \times n$ matrix with (i,j)th element $\mathrm{Cov}(X_i, X_j)$, $1 \le i, j \le n$, and where $\boldsymbol{\gamma}_n \equiv \boldsymbol{\gamma}_n(\theta) = (\mathrm{Cov}(\Psi, X_1), \ldots, \mathrm{Cov}(\Psi, X_n))'$. Here and in the following, we drop θ from population quantities, except when it is important to highlight the dependence on θ and similarly, often drop n from subscript, for simplicity of exposition. Note that by the Pythagorus theorem, the MSPE of the ideal predictor $\tilde{\Psi}_n$ is given by

$$M_{1n}(\theta) = \mathrm{Var}(\Psi) - \boldsymbol{\gamma}_n'\Gamma_n^{-1}\boldsymbol{\gamma}_n.$$

However, exact expressions for the second and the third terms in (1.4) are not easy to write down and both of these terms depend on the particular estimator $\hat{\theta}_n$ is used. In the next section, we describe a resampling method for estimating all three components of the MSPE of the EBLP $\hat{\Psi}_n$.

2.2 Ordinary Bootstrap estimator of the MSPE

Let $\tilde{\theta}_n$ be an estimator of θ based on $X_1, \ldots, X_n$. We shall use $\tilde{\theta}_n$ to produce the bootstrap estimator of the MSPE $M(\theta)$. In principle, one may take $\tilde{\theta}_n = \hat{\theta}_n$, but a different choice of $\hat{\theta}_n$ may be more appropriate in a specific application. The main steps in the ordinary bootstrap estimation procedure are as follows:

I. Generate a bootstrap sample $X_1^*, \ldots, X_n^*$ under the $\theta = \tilde{\theta}_n$. Let θ_n^* denote the bootstrap version of $\hat{\theta}_n$ obtained by replacing $X_1, \ldots, X_n$ by $X_1^*, \ldots, X_n^*$.

II. Compute $\hat{\Psi}_n^* = \tilde{\Psi}_n(\theta_n^*)$ by replacing θ in $\boldsymbol{\lambda}(\theta)$ (cf. (2.1)) by θ_n^*.

III. The bootstrap estimator of $M(\theta)$ is given by

$$\widehat{\text{mspe}}_n^{\text{OR}} = E_* \left(\hat{\Psi}_n^* - \Psi_n^* \right)^2, \tag{2.2}$$

where $\Psi_n^* = \psi(X_1^*, \ldots, X_n^*)$ is the bootstrap version of the predictand Ψ and where E_* denotes conditional expectation given $X_1, \ldots, X_n$.

In practice, evaluation of the conditional expectation is done by the Monte-Carlo method. For this, steps (I)-(III) are repeated a large number (say, B) of times and the resulting bootstrap replicates are combined. Specifically, for each $b = 1, \ldots, B$, one generates the bth resample $X_1^{*b}, \ldots, X_n^{*b}$ under the $\theta = \tilde{\theta}_n$ (independently of the other replicates) and then computes θ_n^{*b}, $\hat{\Psi}_n^{*b} = \tilde{\Psi}_n(\theta_n^{*b})$ and $\Psi_n^{*b} = \psi(X_1^{*b}, \ldots, X_n^{*b})$ based on $X_1^{*b}, \ldots, X_n^{*b}$ as in steps (I)-(III). The Monte-Carlo approximation to $\widehat{\text{mspe}}_n^{\text{OR}}$ is given by

$$\widehat{\text{mspe}}_n^{\text{OR:MC}} = B^{-1} \sum_{b=1}^{B} \left(\hat{\Psi}_n^{*b} - \Psi_n^{*b} \right)^2. \tag{2.3}$$

The following result shows that the ordinary bootstrap estimator of the MSPE is consistent under mild conditions on the underlying time series $\{X_t\}_{t=-\infty}^{\infty}$ and on the estimator sequences $\{\hat{\theta}_n\}_{n \geq 1}$ and $\{\tilde{\theta}_n\}_{n \geq 1}$.

Theorem 2.1: *Let θ_0 denote the true value of the parameter θ and let $\Theta_0 = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \delta_0\}$ for some $\delta_0 \in (0, \infty)$. Suppose that $\tilde{\theta}_n - \theta_0 = o_p(1)$ as $n \to \infty$ and that the following conditions hold:*

(A.1) There exists $\delta \in (0, 1]$ such that
(i) $\sup\{E_\theta \Psi^2 : \theta \in \Theta_0\} < \delta^{-1}$, and
(ii) $\delta < f_\theta(\omega) \leq \delta^{-1}$ for all $\omega \in (-\pi, \pi)$ and $\theta \in \Theta_0$.

(A.2) (i) For each $j \leq 0$, $g_j(\theta)$ is continuous at $\theta = \theta_0$ and $|g_j(\theta)| \leq a_j$ for all $\theta \in \Theta_0$, where $\sum_{j=-\infty}^{0} a_j < \infty$.
(ii) $f_\theta(\cdot)$ is continuous at $\theta = \theta_0$ in the $\|\cdot\|_\infty$-norm.

(A.3) $\sup\{M_{3n}(\theta) : \theta \in \Theta_0\} \to 0$ *as* $n \to \infty$.

Then

$$\widehat{mspe}_n^{\mathrm{OR}} - M_n(\theta_0) \to_p 0 \quad as \quad n \to \infty. \tag{2.4}$$

Conditions (A.1)-(A.3) are local uniformity conditions on various second order population quantities (moments) related to the time series $\{X_i\}_{i=-\infty}^{\infty}$, and essentially requires continuity of the parametric model at $\theta = \theta_0$. Condition (A.1)(i) requires that the second moment of the predictand Ψ be bounded in a neighborhood of the true parameter value θ_0, which would hold if $E_{\theta_0}\Psi^2 < \infty$ and $E_\theta \Psi^2$, as a function of θ, is continuous at $\theta = \theta_0$. Condition (A.1)(ii) is a crucial condition that is used all through the paper. It is used to obtain some bounds on the spectral norm of the matrix $\Gamma_n(\theta)$ and its inverse. This condition is satisfied when $\{X_i\}_{i=-\infty}^{\infty}$ is an ARMA(p, q)-process where all roots of the corresponding characteristic polynomial lie outside the unit circle (cf. Brockwell and Davis (1991)). Next consider (A.2). Continuity of $g_j(\cdot)$ at $\theta = \theta_0$ is tied down to the continuity of the model parametrization at $\theta = \theta_0$. The local uniform summability of $g_j(\theta)$'s can be replaced by requiring finiteness and continuity of the absolute sum $\sum_{j \le 0} |g_j(\theta)|$ on Θ_0, in which case a_j corresponds to $|g_j(\theta_1)|$ for all $j \le 0$, for a common $\theta_1 \in \Theta_0$. Alternatively, it is guaranteed if some standard mixing and moment conditions hold. More specifically, suppose that the process $\{X_i\}_{i=-\infty}^{\infty}$ is strongly mixing with the mixing co-efficient

$$\alpha(n; \theta) \equiv \sup\{|P_\theta(A \cap B) - P_\theta(A)P_\theta(B)| : A \in \mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{-\infty}, j \in \mathbb{Z}\}, \tag{2.5}$$

where $\mathcal{F}_a^b = \sigma\langle X_t : t \in [a, b] \cap \mathbb{Z}\rangle$, $-\infty \le a \le b \le \infty$. Let $\alpha_0(n) = \sup_{\theta \in \Theta_0} \alpha(n; \theta)$, $n \ge 1$. If $E|\Psi|^{2+\delta} < \infty$ and $\sum_{n=1}^{\infty} \alpha_0(n)^{\frac{\delta}{2+\delta}} < \infty$, then (A.2)(i) holds. Condition (A.2)(ii) requires a form of continuity of the parametric model at $\theta = \theta_0$, and is satisfied in many examples, including the class of ARMA (p, q)-models mentioned above. It can be further ascertained if the function $E_\theta X_0 X_j$ is continuous at $\theta = \theta_0$ for each $j \ge 0$, and for some $\delta > 0$, $E|X_1|^{2+\delta} < \infty$ and $\sum_{n=1}^{\infty} \alpha_0(n)^{\frac{\delta}{2+\delta}} < \infty$. Finally, consider (A.3). As pointed out before, the function $M_{3n}(\theta)$ quantifies the effect of replacing the unknown true value of the parameter by the estimator $\hat{\theta}_n$ in the (ideal) BLP, and hence, it critically depends on the properties of the estimator sequence $\{\hat{\theta}_n\}_{n \ge 1}$. Typically, for a sequence of consistent estimators $\{\hat{\theta}_n\}_{n \ge 1}$, $M_{3n}(\theta_0) \to 0$. Condition (A.3) requires the convergence to be uniform in a neighborhood of θ_0. We impose the condition directly on $M_{3n}(\cdot)$ to keep the statement of Theorem 2.1 simple, which only claims consistency of the ordinary bootstrap estimator of the MSPE. A set of sufficient conditions for (A.3) is given in Section 3, where a more precise bound (namely, $O(n^{-1})$) on the order of $M_{3n}(\cdot)$ is obtained.

2.3 Limitations of the ordinary bootstrap estimator

It is easy to see that the ordinary bootstrap estimator of the true MSPE $M_n(\theta_0)$ is equivalent to the plug-in estimator $M_n(\tilde{\theta}_n)$, and has the added advantage that it does not require an explicit expression for the three components $M_{in}(\theta_0),\ i = 1, 2, 3$ (cf. (1.4)). However, as explained earlier, of the three terms in (1.4), only the leading term $M_{1n}(\theta_0) = O(1)$ while the terms $M_{in}(\theta_0),\ i = 2, 3$ are typically of the order $O(n^{-1})$. Therefore, Theorem 2.1 asserts consistency of bootstrap estimator $\widehat{\text{mspe}}_n^{\text{OR}}$ for $M_{1n}(\theta_0)$, the MSPE of the ideal predictor $\tilde{\Psi}_n$, only and fails to capture the effects of estimating the unknown θ_0 by $\hat{\theta}_n$, leading to the terms $M_{in}(\theta_0),\ i = 2, 3$ in the overall MSPE $M_n(\theta_0)$ of the EBLP $\hat{\Psi}_n$. For a better approximation, effects of the terms $M_{in}(\theta_0),\ i = 2, 3$ must be taken into account. In the next section, we describe an implicit bias-correction method based on the bootstrap that achieves this goal.

3 Second order accurate estimation of the MSPE

3.1 The tilting method

We first describe the tilting method in our MSPE estimation problem. The basic idea behind the tilting method is to replace the original estimator $\hat{\theta}_n$ with a suitably *tilted* (or *perturbed*) estimator of θ that annihilates the bias contribution of $\hat{\theta}_n$ to $M_{1n}(\cdot)$, upto the second order accuracy. Suppose that

$$\sum_{j=1}^{p} |M_{1n}^{(j)}(\theta_0)| > \epsilon_0, \tag{3.1}$$

for some $\epsilon_0 > 0$, where for a smooth function $f : I\!R^p \to I\!R$, $f^{(i)}$, $f^{(i,j)}$ and $f^{(i,j,k)}$ denote the first, the second and the third order partial derivatives with respect to the i-th co-ordinate, the (i, j)-th co-ordinates, and the (i, j, k)-th co-ordinates, respectively, $i, j, k = 1, \cdots, p$. Condition (3.1) says that $M_{1n}^{(i)}(\theta_0) \neq 0$ for some i. For notational simplicity, we suppose that $M_{1n}^{(1)}(\theta_0) \neq 0$. Next let $\beta_n \equiv \beta_n(\theta)$ and $\Sigma_n = \Sigma_n(\theta)$ respectively denote the the bias and variance of $\hat{\theta}_n$. We shall also suppose that some consistent estimators $\hat{\beta}_n$ and $\hat{\Sigma}_n$ of β_n and Σ_n, respectively, are available. For example, under mild conditions on $\hat{\theta}_n$ and $\{X_t\}$, such estimators can be generated using the bootstrap method (cf. Lahiri (2003)). Then the *preliminary tilted estimator* of $\boldsymbol{\theta}$ is defined as $\hat{\theta}_n + \mathbf{r}_n$, where $\mathbf{r}_n$ is given by,

$$\mathbf{r}_n = -\left[\sum_{i=1}^{p} M_{1n}^{(i)}(\hat{\theta}_n)\hat{\beta}_{n,i} + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p} M_{1n}^{(i,j)}(\hat{\theta}_n)\hat{\Sigma}_n(i,j)\right]\left\{M_{1n}^{(1)}(\hat{\theta}_n)\right\}^{-1}\mathbf{e}_1, \tag{3.2}$$

where, $\hat{\beta}_{n,i}$ and $\hat{\Sigma}_n(i,j)$ denote the ith component of $\hat{\beta}_n$ and (i,j)th component of $\hat{\Sigma}_n$, respectively, and where the vector $\mathbf{e}_\ell \in I\!R^p$ has one in the ℓth position and zeros elsewhere, $1 \le \ell \le p$. Thus, the preliminary tilted estimator is obtained from the initial estimator $\hat{\theta}_n$ by adding a correction factor to the first component of $\hat{\theta}_n$ only. Note that if, instead of $M_{1n}^{(1)}(\cdot)$, a different partial derivative $M_{1n}^{(i)}(\cdot)$ were nonzero, then we would define the preliminary tilted estimator by replacing the factor $\left\{M_{1n}^{(1)}(\hat{\theta}_n)\right\}^{-1}\mathbf{e}_1$ in (3.2) with $\left\{M_{1n}^{(i)}(\hat{\theta}_n)\right\}^{-1}\mathbf{e}_i$.

To make the MSPE estimator well-defined and to ensure its consistency, we need to modify the preliminary tilted estimator $\hat{\theta}_n + \mathbf{r}_n$. The modifications are needed either if $\hat{\theta}_n + \mathbf{r}_n$ falls outside Θ, in which case $M_n(\hat{\theta}_n + \mathbf{r}_n)$ is not well defined, or if $M_{1i}^{(1)}(\hat{\theta}_n)$ becomes too small, in which case, it scales up the variability of the correction factor $\mathbf{r}_n$. Under appropriate regularity conditions, the probability of getting a preliminary estimator $\hat{\theta}_n + \mathbf{r}_n$ outside Θ or that of getting a value of $M_{1n}^{(1)}(\hat{\theta}_n)$ below the threshold $(1+\log n)^{-2}$ tends to zero rapidly as $n \to \infty$. As a consequence, the perturbed estimator $\check{\theta}_n$ coincides with the preliminary perturbed estimator $\hat{\theta}_n + \mathbf{r}_n$ with high probability.

The *tilted estimator of the MSPE* is now defined as

$$\widehat{\mathrm{MSPE}}_n = M_n(\check{\theta}_n) \tag{3.3}$$

where $\check{\theta}_n$ is the *tilted estimator* of θ, defined by

$$\check{\theta}_n = \begin{cases} \hat{\theta}_n + \mathbf{r}_n & \text{if } \hat{\theta}_n + \mathbf{r}_n \in \Theta \text{ and } |M_{1n}^{(1)}(\hat{\theta}_n)|^{-1} \le (1+\log n)^2 \\ \hat{\theta}_n & \text{otherwise.} \end{cases} \tag{3.4}$$

Although an explicit expression for the function $M_n(\cdot)$ is typically unknown, it is not difficult to see that the tilted estimator $\widehat{\mathrm{MSPE}}_n$ is equivalently given by (2.2) with $\tilde{\theta}_n = \check{\theta}_n$; The latter can be computed using the algorithm given in Section 2.2.

In the next section, we state the regularity conditions and show that the tilted estimator of the MSPE in (3.3) achieves second order bias accuracy.

3.2 Theoretical properties

As before, let θ_0 denote the true value of the parameter θ and let $\Theta_0 = \{\theta \in \Theta : \|\theta - \theta_0\| \le \delta_0\}$ denote a open neighborhood of θ_0. Let P_θ and E_θ denote the probability and expectation under θ. For notational simplicity, we set $P_{\theta_0} = P$ and $E_{\theta_0} = E$. For $j \in Z\!\!\!Z$, define

$$g_j(\theta) = E_\theta[\psi(X_1, \ldots, X_k)X_j], \quad \theta \in \Theta.$$

Note that $g_j(\theta)$ is the covariance between X_{n+j} and $\Psi = \psi(X_{n+1}, \ldots, X_{n+k})$ under θ, which decreases to zero as $j \to -\infty$ under suitable weak dependence and moment conditions on

$\{X_i\}_{i=-\infty}^{\infty}$. Let $\Delta\boldsymbol{\lambda}_n(\theta)$ be the $p \times n$ matrix, with ith column given by the $p \times 1$ vector of partial derivatives of the ith component of $\boldsymbol{\lambda}_n(\theta) = \gamma_n(\theta)\Gamma_n(\theta)^{-1}$. Define

$$\mu_{2n}(\theta) \equiv nE_\theta\Big([\hat{\theta}_n - \theta]'\Delta\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n\{\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n - \Psi\}\Big)$$
$$\mu_{3n}(\theta) \equiv E_\theta\Big(n^{1/2}[\hat{\theta}_n - \theta]'\Delta\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n\Big)^2,$$

which respectively give approximations to the functions $M_{2n}(\theta)$ and $M_{3n}(\theta)$, upto an error of order $o(n^{-1})$.

Conditions:

(C.1) Suppose that there exists a $\delta \in (0, \infty)$ such that

$$\liminf_{n\to\infty} M_{1n}^{(1)}(\theta_0) \geq \delta.$$

(C.2) Suppose that there exists $\kappa, c_0 \in (0, \infty)$ such that for all $\theta \in \Theta$:

- (i) $E_\theta \Psi^2 < c_0$,
- (ii) $E_\theta |X_1|^{4+\kappa} < c_0$,
- (iii) $\limsup_{n\to\infty} E_\theta \Big\{\sqrt{n}\|\hat{\theta}_n - \theta\|\Big\}^8 < c_0$.

(C.3) Suppose that $g_j(\theta)$ and f_θ are twice differentiable on Θ, and that there exist a constant $c_1 \in (0, \infty)$ and a sequence $\{a_n\}_{n\geq 1} \subset (0, \infty)$ with $\sum_{n-1}^{\infty} a_j < \infty$ such that for all $k, l \in \{1, \ldots, p\}$,

- (i) $\max\{g_j(\theta), |g_j^{(k)}(\theta)|, |g_j^{(k,l)}(\theta)|\} < a_j$ for all $\theta \in \Theta$,
- (ii) $\max\{\|f_\theta\|_\infty, \|f_\theta^{-1}\|_\infty, \|f_\theta^{(k)}\|_\infty, \|f_\theta^{(k,l)}\|_\infty\} < c_1$ for all $|a| \leq 2$ and for all $\theta \in \Theta$, and
- (iii) $\|f_\theta^{(k,l)} - f_{\theta_0}^{(k,l)}\| \leq c_1\|\theta - \theta_0\|^\delta$, for all $\theta \in \Theta_0$ for some $\delta > 0$.

(C.4) Suppose that there exists a $c_2 \in (0, \infty)$ such that $\sup\{|\mu_{kn}(\theta)| : \theta \in \Theta\} < c_2$ for all $n \geq c_2$, and $\mu_{kn}(\cdot)$ is equi-continuous at $\theta = \theta_0$, $k = 2, 3$.

(C.5) Suppose that $\beta_n(\theta) = n^{-1}\beta_0(\theta) + o(n^{-1})$ and $\Sigma_n(\theta) = n^{-1}\Sigma_0 + o(n^{-1})$ uniformly in $\theta \in \Theta$ and $\Delta_0 \equiv \sup\{\|\beta_0\| + \|\Sigma_0(\theta)\| : \theta \in \Theta\} < \infty$.

Condition (C.1) is a specialized version of (3.1) for the given formula for the correction factor $\mathbf{r}_n$, which says that the function $M_{1n}(\cdot)$ has a non-zero derivative along one of the directions $i \in \{1, \ldots, p\}$ at the true value θ_0, and is typically satisfied in most applications.

See the discussion following (3.1) for implications and alternative versions of this. Condition (C.2)(i) is needed to make $M_{1n}(\cdot)$ well-defined while Conditions (C.2)(ii) and (iii) are used to establish exact orders of the functions $M_{kn}(\cdot)$ for $k = 2, 3$ (cf. Lemma 6.2 below). Condition (C.3) is a smoothness condition on the spectral density of the process $\{X_t\}$ and on the cross-covariances $g_j(\theta) = \mathrm{Cov}_\theta(\Psi, X_j)$, which would hold if the underlying model-parametrization is suitably smooth. The same comment applies to Condition (C.4), which requires boundedness and equi-continuity of the approximating functions μ_{kn}, $k = 2, 3$. Finally, Condition (C.5) is a condition on the bias and the variance of the estimator sequence $\{\hat{\theta}_n\}_{n\geq 1}$. We have decided to state Conditions (C.4) and (C.5) in terms of the original sequence $\{\hat{\theta}_n\}_{n\geq 1}$ to allow for generality. For a specific choice of $\hat{\theta}_n$, these conditions have to be checked directly. To indicate the type of arguments one would need to verify (suitable variants) of these conditions, consider the class of estimator sequences $\{\hat{\theta}_n\}_{n\geq 1}$ that admit a representation of the form:

$$\hat{\theta}_n - \theta = \frac{\beta_0(\theta)}{n} + n^{-1}\sum_{i=1}^{n} \xi_i + R_n \tag{3.5}$$

for some function $\beta_0(\cdot) : \Theta \to I\!R^p$, zero mean random vectors $\xi_i \in \sigma\langle X_i \rangle$, $i \geq 1$ and a remainder term R_n. Suppose that there exist constants $\delta, c_3 \in (0, \infty)$ and a sequence $\{d_n\}_{n\geq 1}$ satisfying $d_n = o(n^{-1/2})$ such that $\|\beta(\theta)\| < c_3$, $E_\theta\|\xi_i\|^8 < c_3$, $E_\theta\|d_n^{-1} R_n\|^8 < c_3$ and $\sum_{n=1}^{\infty} n^3 \alpha(n;\theta)^{\frac{\delta}{8+\delta}} < c_3$ for all $\theta \in \Theta$. Then, it is easy to check that Conditions (C.2)(iii) and (C.5) hold. Under (3.5), it can be shown that a variant of Condition (C.4) holds where the factor $n^{1/2}(\hat{\theta}_n - \theta)$ in the functions μ_{kn} are replaced by the leading to terms from (3.5). For example, for $k = 3$, it can be shown that under (C.3),

$$\sup\Big\{\Big|\mu_{3n}(\theta) - \tilde{\mu}_{3n}(\theta) : \theta \in \Theta\Big\} = o(1). \tag{3.6}$$

where

$$\tilde{\mu}_{3n}(\theta) \equiv E_\theta\Big(\Big[n^{-1/2}\sum_{i=1}^{n} \xi_i\Big]' \Delta\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n\Big)^2.$$

As a result, one can use $\tilde{\mu}_{3n}(\theta)$ in place of $\mu_{3n}(\theta)$ as an approximation to $M_{3n}(\theta)$ to establish Theorem 3.1 (retracing the steps given in Section 6). Note that the equi-continuity of $\tilde{\mu}_{3n}(\theta)$ at $\theta = \theta_0$ can now be proved under a continuity condition on the individual lag-covariance functions $E_\theta[X_1, \xi_1][X_{k+1}, \xi_{k+1}]'$, $k \geq 0$ (as functions of θ) as in Condition (A.2) and the discussion following the statement of Theorem 2.1. We give a proof of (3.6) in Section 6. A similar treatment is possible also for the term $\mu_{2n}(\theta)$. Hence, it follows that for an estimator sequence $\{\hat{\theta}_n\}_{n\geq 1}$ satisfying (3.5), Conditions (C.2) - (C.5) hold under mild moment conditions on the variables X_t's and ξ_t's and under mild weak dependence conditions on the underlying process.

With this, we are now ready to state the main result of this section.

Theorem 3.1: *Suppose that Conditions (C.1) - (C.5) hold. Then*

$$E\Big(\widehat{\text{MSPE}}_n - M_n(\theta_0)\Big) = o(n^{-1}) \tag{3.7}$$

$$\text{Var}(\widehat{\text{MSPE}}_n - M_n(\theta_0)\Big) = O(n^{-1}). \tag{3.8}$$

Theorem 3.1 shows that under suitable regularity conditions, the tilted MSPE estimator attains second order bias accuracy. Further, the variance of the tilted estimator continues to be of the same order as the untilted (naive) MSPE estimator, and is guaranteed to be non-negative. Thus, the tilted MSPE estimator may be preferred over the ordinary MSPE estimator that fails to capture the effects of parameter estimation in the EBLP on the overall MSPE. In the next section, we describe some important issues related to the implementation of the titling method in practice.

4 Practical implementation based on the bootstrap

Note that the tilting method described above involves computing the functions $M_{in}(\cdot)$, $i = 1, 2, 3$, and its first and second order partial derivatives, for which explicit expressions are not always available. In this section, we develop bootstrap based approximations to these quantities, so that the tilted MSPE estimator can be used in practice, without any analytical derivations. To that end, first we define a bootstrap-based approximation to the function $M_{1n}(\cdot)$ at a given value $\theta = \theta_1$ (which may depend on the data). The steps are similar to those used for generating the Monte-carlo approximation to $\widehat{\text{mspe}}_n^{\text{OR}}$ in (2.3). Specifically, for $b = 1, \cdots, B$,
(i) generate bootstrap samples $(X_1^{*b}, \cdots, X_{n+k}^{*b})$ under θ_1,
(ii) compute $\tilde{\Psi}_n^{*b}$ and Ψ^{*b} by replacing $X_1, \ldots, X_n$ with $X_1^{*b}, \cdots, X_{n+k}^{*b}$. The Monte-carlo approximation to $M_{1n}(\theta_1)$ is given by

$$M_{1n}^*(\theta_1) = B^{-1} \sum_{b=1}^{B} (\tilde{\Psi}_n^{*b} - \Psi^{*b})^2. \tag{4.1}$$

Next we construct estimates of the partial derivatives of the function $M_{1n}(\cdot)$ for computing the correction factor $\mathbf{r}_n$. To motivate the construction, first consider a smooth function

$g: I\!R \to I\!R$. Then, for any $x \in I\!R$, using Taylor series expansion,

$$g(x+\epsilon) - g(x-\epsilon) = 2\epsilon g^{'}(x) + o(\epsilon),$$

as $\epsilon \to 0$, where $g^{'}(x)$ denotes the derivative of $g(x)$ at x. Hence we can use the scaled difference $(2\epsilon)^{-1}\{g(x+\epsilon) - g(x-\epsilon)\}$ as an approximation to $g^{'}(x)$ for small values of $\epsilon > 0$. Relying on this fact, we can now define suitable bootstrap approximations to the first order partial derivatives of $M_{1n}(\cdot)$ at $\hat{\theta}_n$. Let $\{a_n\}_{n\geq 1}$ be a sequence of positive real numbers converging to zero. Then, with M^*_{1n} as in (4.1), we define the bootstrap approximation to the first order partial derivatives as,

$$M^{*(j)}_{1n}(\hat{\theta}_n) = (2a_n)^{-1}\left[M^*_{1n}(\hat{\theta}_n + a_n\mathbf{e}_j) - M^*_{1n}(\hat{\theta}_n - a_n\mathbf{e}_j)\right], \tag{4.2}$$

$j = 1, \cdots, p$. Similarly, we can define the bootstrap approximations to the second order partial derivatives as:

$$\begin{aligned} M^{*(j,j)}_{1n}(\hat{\theta}_n) &= a_n^{-2}\left[M^*_{1n}(\hat{\theta}_n + a_n\mathbf{e}_j) + M^*_{1n}(\hat{\theta}_n - a_n\mathbf{e}_j) - 2M^*_{1n}(\hat{\theta}_n)\right], \quad 1 \leq j \leq p, \\ M^{*(i,j)}_{1n}(\hat{\theta}_n) &= 2a_n^{-2}\left[\left\{M^*_{1n}(\hat{\theta}_n + a_n\mathbf{e}_{i,j}) + M^*_{1n}(\hat{\theta}_n - a_n\mathbf{e}_{i,j}) - 2M^*_{1n}(\hat{\theta}_n)\right\}\right. \\ &\qquad \left. - a_n^2\left\{M^{*(i,i)}_{1n}(\hat{\theta}_n) + M^{*(j,j)}_{1n}(\hat{\theta}_n)\right\}\right], \quad 1 \leq i \neq j \leq p. \end{aligned} \tag{4.3}$$

where, $\mathbf{e}_{i,j} = \mathbf{e}_i + \mathbf{e}_j$.

Then, we have the following result on the accuracy of the bootstrap estimates of the partial derivatives:

Proposition 4.1 *Suppose Conditions (C.2) and (C.3) hold. Then,*

$$E_*\left|M^{*(j)}_{1n}(\hat{\theta}_n) - M^{(j)}_{1n}(\hat{\theta}_n)\right|^2 = O(B^{-1}a_n^{-2} + a_n^2) \quad \textit{almost surely}$$

for all $1 \leq j \leq p$ *and*

$$E_*\left|M^{*(i,j)}_{1n}(\hat{\theta}_n) - M^{(i,j)}_{1n}(\hat{\theta}_n)\right|^2 = O(B^{-1}a_n^{-4} + a_n^2) \quad \textit{almost surely}$$

for all $1 \leq i, j \leq p$.

Thus, by choosing a_n small and then choosing the number of bootstrap replicates B suitably large, we can generate accurate approximations to the first and second order partial derivatives of the function $M_{1n}(\cdot)$. Analytical derivations of the partial derivatives, therefore, can be completely bypassed by using the bootstrap (and hence, necessary computing resources).

Next, we define the bootstrap estimators of the bias and variance of $\hat{\theta}_n$ by,

$$\beta_n^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_n^{*b} - \hat{\theta}_n,$$

$$\Sigma_n^* = \left\{\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_n^{*b}(\hat{\theta}_n^{*b})'\right\} - \left(\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_n^{*b}\right)\left(\frac{1}{B}\sum_{b=1}^{B}\hat{\theta}_n^{*b}\right)' \tag{4.4}$$

respectively, where $\hat{\theta}_n^{*b}$ denote the bth bootstrap replicate of $\hat{\theta}_n$, obtained by replacing $X_1, \ldots, X_n$ with $X_1^{*b}, \cdots, X_n^{*b}$ and $\{(X_1^{*b}, \cdots, X_n^{*b}) : b = 1, \ldots, B\}$ are independent bootstrap replicates under $\theta = \hat{\theta}_n$.

Proposition 4.2 *Suppose Conditions (C.2), (C.3) and (C.5) hold. Then,*

$$E_*\left\|\beta_n^* - \beta_n(\hat{\theta}_n)\right\|^2 = O(B^{-1}n^{-2}) + o(n^{-2}) \quad \textit{almost surely}$$

and

$$E_*\left\|\Sigma_n^* - \Sigma_n(\hat{\theta}_n)\right\|^2 = O(B^{-1}n^{-2}) + o(n^{-2}) \quad \textit{almost surely}$$

Combining (4.2), (4.3) and (4.4) , we now define the bootstrap based correction factor as

$$\mathbf{r}_n^* = -\left[\sum_{i=1}^{p} M_{1n}^{*(i)}(\hat{\theta}_n)\beta_{n,i}^* + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}\{M_{1n}^{*(i,j)}(\hat{\theta}_n)\}\{\Sigma_n^*(i,j)\}\right]\frac{\mathbf{e}_1}{\left\{M_{1n}^{*(1)}(\hat{\theta}_n)\right\}}, \tag{4.5}$$

where $\beta_{n,i}^*$ and $\Sigma_n^*(i,j)$ respectively denote the ith component of β_n^* and the (i,j)th element of Σ_n^*, $1 \le i,j \le p$. The bootstrap-based bias-corrected MSPE estimate is given by

$$\widehat{\text{MSPE}}_n^* = \widehat{\text{mspe}}_n^{\text{OR:MC}}(\check{\theta}_n^*) \tag{4.6}$$

where $\widehat{\text{mspe}}_n^{\text{OR:MC}}(\check{\theta}_n^*)$ is defined by (2.3) with $\tilde{\theta}_n = \check{\theta}_n^*$, and $\check{\theta}_n^*$ is defined by replacing $\mathbf{r}_n$ and $M_{1n}^{(1)}(\hat{\theta}_n)$ in (3.4) by $\mathbf{r}_n^*$ and $M_{1n}^{*(1)}(\hat{\theta}_n)$, respectively.

In view of Theorem 3.1 and Propositions 4.1 and 4.2, $\widehat{\text{MSPE}}_n^*$ gives an accurate approximation to the bias-corrected estimator of the MSPE that can be evaluated without any analytical work, provided Conditions (C.1)-(C.5) hold. However, finite sample performance of the MSPE estimator depends on the choice of different factors, such as a_n, B, etc. In the next section, we explore these issues further through a simulation study.

5 Simulation study

For the simulation study, we consider one-step-ahead best linear prediction, i.e., we take the predictand Ψ to be X_{n+1}. We shall consider the the following time series models.

Model 1: In the linear autoregressive model of order 2, AR(2),

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \epsilon_t, \tag{5.1}$$

for $t \in \mathbb{Z}$, where $\{\epsilon_t\}$ are independent and identically distributed $N(0, \sigma^2)$. Values of ϕ_1, ϕ_2 are chosen to be $(0.2, 0.5)$ and the value of σ^2 is 4.
Model 2: In the linear autoregressive moving-average model of order (1,1), ARMA(1,1),

$$X_t = \phi_1 X_{t-1} + \epsilon_t + \psi_1 \epsilon_{t-1}, \tag{5.2}$$

for $t \in \mathbb{Z}$, where $\{\epsilon_t\}$ are independent and identically distributed $N(0, \sigma^2)$. Here we take the parameter values to be $\phi_1 = 0.2, \psi_1 = 0.5$ and we take $\sigma^2 = 4$.
In this simulation study, we will perturb the estimator in the direction of σ^2. In implementing the method, we use $B = 1000$ bootstrap samples to estimate the bias, variances and for all other approximations. All simulation results are based on $N = 500$ replications. The simulations are done for $n = 50, 120$ and 500. Table 1 reports the empirical measures of bias and root mean squared error (RMSE) for both bias-corrected and not bias-corrected estimators $\widehat{\text{MSPE}}_n^*$ and $\widehat{\text{mspe}}_n^{\text{OR:MC}}$ of MSPE for three different values of n. The bias and mean squared error (MSE) are estimated empirically by taking the average over the replicates of the bias and MSE for each dataset. From Table 1 we can see that the bias correction method gives us significantly better results for different values of n under the models (5.1)and(5.2). However, it is worth mentioning that due to the bias correction, the RMSE's of the bias-corrected estimators are seemed to be slightly higher than the not bias-corrected estimators. This is expected, as the randomness in the various approximation steps in the construction of the bias-corrected estimators adds to its total variability. Boxplots for RMSE's of the two estimators of MSPE over $N = 500$ simulations under different models are presented in Figure 1. The boxplots also support the conclusions obtained from Table 1. In these boxplots we can see that due to the bias correction the RMSE's of the tilted estimators seem to be higher than the unperturbed estimators.

6 Proofs

For a $l \times l$ matrix A, let $\|A\| = \sup\{\|Ax\| : x \in \mathbb{R}^l, \|x\| = 1\}$ denote the spectral norm, where $1 \le l \le \infty$ and where $\|\cdot\|$ denotes the ℓ^2 norm on $\mathbb{R}^2$. Let $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. For $a = (a_1, \ldots, a_p)' \in \mathbb{Z}_+^p$, let $|a| = |a_1| + \ldots + |a_p|$, $a! = \prod_{i=1}^p a_i!$ and $D^a = D_1^{a_1} \ldots D_p^{a_p}$, where D_j denotes the partial derivative w.r.t. the jth co-ordinate, $1 \le j \le p$. Let $C, C(\cdot)$ denote generic constants with values in $(0,\infty)$ that depend on their arguments, if any, but not on n. Unless otherwise specified, limits in order symbols are taken by letting $n \to \infty$.

6.1 Auxiliary Lemmas

Lemma 6.1 *Suppose that $g_j(\theta)$ and f_θ are twice differentiable on Θ, and that there exists a constant $C_0 > 0$ such that*

(i) $\sum_{j\le 0} |D^a g_j(\theta)|^2 < C_0$ *for all* $\theta \in \Theta$,

(ii) $\|f_\theta\|_\infty + \|f_\theta^{-1}\|_\infty + \|D^a f_\theta\|_\infty < C_0$ *for all* $|a| \le 2$ *and for all* $\theta \in \Theta$, *and*

(iii) $\|D^a f_\theta - D^a f_{\theta_0}\| \le C_0 \|\theta - \theta_0\|^\delta$, *for all* $\theta \in \Theta_0$ *for some* $\delta > 0$.

Then, there exists a constant $C_1 \in (0,\infty)$ *such that* $|M_{1n}^{(j)}(\theta)| + |M_{1n}^{(i,j)}(\theta)| < C_1$ *for all* $\theta \in \Theta$ *and for all* $n \ge C_1$, *where* $1 \le i,j \le p$. *Further,* $\sum_{1\le i,j\le p} |M_{1n}^{(i,j)}(\theta_0) - M_{1n}^{(i,j)}(\theta)| \le C_1 \|\theta_0 - \theta\|^\delta$ *for all* $\theta \in \Theta_0$ *and for all* $n \ge C_1$.

Proof: It is easy to check that for $\theta_1, \theta_2 \in \Theta$,

$$
\begin{aligned}
&\gamma_n(\theta_1)'\Gamma_n(\theta_1)^{-1}\gamma_n(\theta_1) - \gamma_n(\theta_2)'\Gamma_n(\theta_2)^{-1}\gamma_n(\theta_2) \\
=& \Big(\gamma_n(\theta_1) - \gamma_n(\theta_2)\Big)'\Gamma_n(\theta_1)^{-1}\gamma_n(\theta_1) + \gamma_n(\theta_2)\Gamma_n(\theta_1)^{-1}\Big(\gamma_n(\theta_1) - \gamma_n(\theta_2)\Big) \\
&+\gamma_n(\theta_2)'\Gamma_n(\theta_2)^{-1}\Big[\Gamma_n(\theta_2) - \Gamma_n(\theta_1)\Big]\Gamma_n(\theta_1)^{-1}\gamma_n(\theta_2),
\end{aligned}
$$

which, in view of conditions (i),(ii) and eqrefP.3, readily implies that

$$
\begin{aligned}
D_j M_{1n}(\theta) = & [D_j\gamma_n(\theta)]'\Gamma_n(\theta)^{-1}\gamma_n(\theta) + \gamma_n(\theta)'\Gamma_n(\theta)^{-1}[D_j\gamma_n(\theta)] \\
&-\gamma_n(\theta)'\Gamma_n(\theta)^{-1}[D_j\Gamma_n(\theta)]\Gamma_n(\theta)^{-1}\gamma_n(\theta).
\end{aligned}
$$

Next using similar arguments for the second derivation (which is now given by nine terms), one can complete the proof of the lemma. We omit the details.

Lemma 6.2: *Suppose that there exists* $\kappa, C \in (0,\infty)$ *such that*

(i) $E_\theta|X_1|^{4+\kappa} < C$,

(ii) $\limsup_{n\to\infty} E_\theta\|\hat{\theta}_n - \theta\|^8 < C$

(iii) $\limsup_{n\to\infty} \|\gamma_n(\theta)\| < C$, *and*

(iv) $\|f_\theta^{-1}\|_\infty < C$

for all $\theta \in \Theta$. *Then,* $\sup_{\theta\in\Theta}\Big[\Big|M_{3n}(\theta) - n^{-1}\mu_{3n}(\theta)\Big| + \Big|M_{2n}(\theta) - n^{-1}\mu_{2n}(\theta)\Big|\Big] = o(n^{-1})$.

Proof: Note that on the set $A_n \equiv \{\|\theta - \hat\theta_n\| \le \epsilon\}$,

$$
\begin{aligned}
&[\boldsymbol{\lambda}_n(\hat\theta_n) - \boldsymbol{\lambda}(\theta)]\mathbf{X}_n \\
&= [\hat\theta_n - \theta]'[\Delta\boldsymbol{\lambda}_n(\theta)]\mathbf{X}_n + \sum_{|a|=2}[\hat\theta_n - \theta]^a D^a\boldsymbol{\lambda}(\theta_1)\mathbf{X}_n/a!
\end{aligned}
$$

where θ_1 is a point in A_n. Hence,

$$
\begin{aligned}
&\Big|M_{3n}(\theta) - E_\theta\Big([\hat\theta_n - \theta]'\Delta\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n\Big)^2\Big| \\
&\le C(p)\sup\{|D^a\boldsymbol{\lambda}(t)|\|\|t-\theta\| \le \epsilon, |a| = 2\}\dot E_\theta\|\hat\theta_n - \theta\|^4\|\mathbf{X}_n\|^2\mathbb{1}(A_n) \\
&\qquad\qquad + E_\theta\Big([|\boldsymbol{\lambda}_n(\hat\theta_n)\mathbf{X}_n| + |\boldsymbol{\lambda}_n(\theta)\mathbf{X}_n|]\cdot\mathbb{1}(\|\hat\theta_n - \theta\| \ge \delta)\Big)^2 \\
&\equiv I_1 + I_2 + I_3, \quad \text{say.}
\end{aligned}
$$

First consider I_2. Let $\mathbf{X}_{n,i} = (X_{1,i},\ldots,X_{n,i})'$, $i = 1,2$ where $X_{j,1} = X_j\mathbb{1}(|X_j| \le c_n)$ and $X_{j,2} = X_j - X_{j,1}$, $1 \le j \le n$, where $c_n = n^{1/2-\kappa/16}$. By (iii) and (iv), there exists $c_0 \in (0,\infty)$ such that $\|\boldsymbol{\lambda}_n(\theta)\|^2 = \gamma_n(\theta)'\Gamma(\theta)^{-2}\gamma_n(\theta) < c_0^2$ for all $\theta \in \Theta$, for n large. Hence, we have, for any $\epsilon > 0$,

$$
\begin{aligned}
&E_\theta\Big([\boldsymbol{\lambda}_n(\hat\theta_n)\mathbf{X}_n]\cdot\mathbb{1}(\|\hat\theta_n - \theta\| \ge \epsilon)\Big)^2 \\
&\le 2E_\theta\Big([\boldsymbol{\lambda}_n(\hat\theta_n)'\mathbf{X}_{n,1}]\cdot\mathbb{1}(A_n^c)\Big)^2 + 2E_\theta\Big(\boldsymbol{\lambda}_n(\hat\theta_n)'\mathbf{X}_{n,2}\Big)^2 \\
&\le 2c_0^2\Big[\Big|E_\theta\Big(\sum_{i=1}^n[X_{i,1}^2 - E_\theta X_{1,1}^2]\mathbb{1}(A_n^c)\Big)\Big| + nE_\theta X_{1,1}^2P_\theta(A_n^c)\Big] + 2c_0^2E_\theta\|\mathbf{X}_{n,2}\|^2 \\
&\le Cc_0^2\Big[\Big(n\sum_{i=1}^\infty|\mathrm{Cov}(X_{1,1}^2, X_{i,1}^2)|\Big)^{1/2}\Big(P_\theta(A_n^c)\Big)^{1/2} + nE_\theta X_1^2P_\theta(A_n^c) \\
&\qquad\qquad + E_\theta X_1^2\mathbb{1}(|X_1| > c_n)\Big] \\
&\le Cc_0^2n^{-1-\epsilon}
\end{aligned}
$$

for some $\epsilon = \epsilon(\kappa) > 0$.

By similar arguments, $I_3 \le Cc_0^2 n^{-1-\epsilon}$. Also,

$$\begin{aligned}
I_1 &\le CE_\theta \|\hat{\theta}_n - \theta\|^4 \|bfX_n\|^2 1\!\!1(A_n) \\
&\le E_\theta\Big(\|\hat{\theta}_n - \theta\|^4 \Big[\sum_{i=1}^n (X_{i,1}^2 - E_\theta X_{i,1}^2) + nE_\theta X_{1,1}^2 + \sum_{i=1}^n (X_{2,1}^2 \Big] 1\!\!1(A_n) \Big) \\
&\le CE_\theta \Big(\|\hat{\theta}_n - \theta\|^8 1\!\!1(A_n) \Big)^{1/2} \Big(n \sum_{i=1}^\infty |\mathrm{Cov}(X_{1,1}^2, X_{i,1}^2)| \Big)^{1/2} \\
&\quad + CnE_\theta X_1^2 E_\theta \|\hat{\theta}_n - \theta\|^4 1\!\!1(A_n) + Cn(E_\theta X_{2,1}^4)^{1/2} (E_\theta \|\hat{\theta}_n - \theta\|^8 1\!\!1(A_n))^{1/2} \\
&\le Cc_0^2 n^{-1-\epsilon}
\end{aligned}$$

for some $\epsilon = \epsilon(\kappa) > 0$.

The proof of the second relation follows by repeating the same arguments, and therefore, it is omitted.

Lemma 6.3: For $j \ge 1$ and $1 \le k \le 4$, let ξ_{kj} be a $\sigma\langle X_j \rangle$-measurable zero-mean random variable such that for some $\delta, c_1 \in (0, \infty)$, $E_\theta |\xi_{kj}|^{4+\delta} < c_1$ for all j, k and $\sum_{n=1}^\infty n^3 \alpha(n;\theta)^{\frac{\delta}{4+\delta}} < c_1$ for all $\theta \in \Theta$. Let $\{e_{kjn} : 1 \le j \le n\}_{n \ge 1} \subset I\!\!R$ be such that $\sum_{j=1}^n e_{kjn}^2 = O(1)$ for $1 \le k \le 4$. Then there exists a constant C_1 (depending on c_1, but not on θ) such that

$$\limsup_{n\to\infty} \Big\{ \Big| E_\theta \Big[\Big(\sum_{i=1}^n \xi_{1i} \Big) \prod_{k=2}^3 \Big(\sum_{i=1}^n e_{kjn}\xi_{ki} \Big) \Big] \Big| + E_\theta \Big[\prod_{k=1}^4 \Big(\sum_{i=1}^n e_{kjn} \xi_{ki} \Big) \Big] \Big\} < C_1$$

for all $\theta \in \Theta$.

Proof: We shall give a proof of the bound on the second term only; the proof of the bound on the first term is similar (and somewhat simpler). Clearly, for any $1 \le k, l \le 4$,

$$\begin{aligned}
& E_\theta \Big(\sum_{i=1}^n e_{kin} \xi_{ki} \Big) \Big(\sum_{j=1}^n e_{ljn} \xi_{lj} \Big) \\
& \sum_{|m|=0}^{n-1} \sum_{\{(i,j):\, i-j=m, 1\le i,j \le n\}} e_{kin} e_{ljn} E_\theta \xi_{ki} \xi_{lj} \\
\le & \sum_{|m|=0}^{n-1} \sum_{\{(i,j):\, i-j=m, 1\le i,j \le n\}} |e_{kin} e_{ljn}| \Big(E_\theta |\xi_{ki}|^{2+\delta} \Big)^{\frac{1}{2+\delta}} \Big(E_\theta |\xi_{lj}|^{2+\delta} \Big)^{\frac{1}{2+\delta}} \alpha(|m|;\theta)^{\frac{\delta}{2+\delta}} \\
\le & \sum_{|m|=0}^{n-1} \Big[\sum_{i=1}^n e_{kin}^2 \Big]^{1/2} \Big[\sum_{i=1}^n e_{lin}^2 \Big]^{1/2} C(c_1, \delta) \alpha(|m|;\theta)^{\frac{\delta}{2+\delta}} \\
\le & C(c_1, \delta) \sum_{m=0}^{n-1} \alpha(m;\theta)^{\frac{\delta}{2+\delta}}.
\end{aligned}$$

Let $\mathcal{K}_4(V_1, V_2, V_3, V_4)$ denote the fourth order (mixed) cumulant of a set of random variables V_1, V_2, V_3, V_4 under θ, defined by

$$\mathcal{K}_4(V_1, V_2, V_3, V_4; \theta) = \frac{\partial}{\partial v_1}\frac{\partial}{\partial v_2}\frac{\partial}{\partial v_3}\frac{\partial}{\partial v_4} E_\theta \exp(\sqrt{-1}[v_1 V_1 + v_2 V_2 + v_3 V_3 + v_4 V_4])\Big|_{v_1=\ldots=v_4=0}.$$

Then, by using multi-linearity of $\mathcal{K}_4(\cdot)$, it follows that

$$\begin{aligned}
&E_\theta\Big[\prod_{k=1}^{4}\Big(\sum_{i=1}^{n} e_{kjn}\xi_{ki}\Big)\Big] \\
&\leq \Big|\mathcal{K}_4(\sum_{i=1}^{n} e_{1jn}\xi_{1i}, \ldots, \sum_{i=1}^{n} e_{4jn}\xi_{4i})\Big| + \sum_{I\subset\{1,2,3,4\},|I|=2}\Big|\mathcal{K}_2(\sum_{i=1}^{n} e_{kjn}\xi_{ki},\ k\in I)(\sum_{i=1}^{n} e_{kjn}\xi_{ki},\ k\in I^c)\Big|.
\end{aligned}$$

Note that

$$\begin{aligned}
&|\mathcal{K}_2(\sum_{i=1}^{n} e_{kjn}\xi_{1i},\ k\in I)| \\
&\leq \prod_{k\in I}\Big[\mathrm{Var}\Big(\sum_{i=1}^{n} e_{kin}\xi_{ki}\Big)\Big]^{1/2} \\
&\leq \prod_{k\in I}\Big[\sum_{i=1}^{n} e_{kin}^2 E_\theta(\xi_{ki})^2 + 2\sum_{j=1}^{n-1}\Big(\sum_{i=1}^{n-j} e_{kin}^2\Big)^{1/2}\Big(\sum_{i=j+1}^{n} e_{kin}^2\Big)^{1/2}|\mathrm{Cov}(\xi_{ki}, \xi_{k(i+j)})|\Big]^{1/2} \\
&= O(1) \quad \text{uniformly in} \quad \theta\in\Theta.
\end{aligned}$$

Next writing $\breve{e}_{in} = \max\{|e_{kin}| : k = 1, 2, 3, 4\}$, $1 \leq i \leq n$, and writing $\sum_b$ for the sum over all $i_1, \ldots, i_4 \in \{1, \ldots, n\}$ with maximal gap b, $0 \le b \le n-1$, we have

$$
\begin{aligned}
&\Big|\mathcal{K}_4(\sum_{i=1}^{n} e_{1jn}\xi_{1i}, \ldots, \sum_{i=1}^{n} e_{4jn}\xi_{4i})\Big| \\
&\le C \sum_{b=0}^{n-1} \sum_{b} \prod_{k=1}^{4} |e_{ki_kn}||\mathcal{K}_4(\xi_{1i_1}, \ldots, \xi_{4i_4})| \\
&\le C \sum_{b=0}^{n-1} \sum_{b} \prod_{k=1}^{4} |e_{ki_kn}| c_1^{\frac{4}{4+\delta}} \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \\
&\le C(c_1,\delta) \sum_{b=0}^{n-1} \Big[\sum_{i=1}^{n} \check{e}_{in} \Big\{ \sum_{|i_k - i| \le b, k=1,2,3} \prod_{k=1}^{3} \check{e}_{i_kn} \Big\} \Big] \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \\
&\le C(c_1,\delta) \sum_{b=0}^{n-1} \Big[\sum_{i=1}^{n} \check{e}_{in} \prod_{k=1}^{3} b^{1/2} \Big(\sum_{|i_k - i| \le b} \check{e}_{i_kn}^2 \Big)^{1/2} \Big] \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \\
&\le C(c_1,\delta) \sum_{b=0}^{n-1} b^{3/2} \Big[\Big\{ \sum_{i=1}^{n} \check{e}_{in}^2 \Big\}^{1/2} \Big\{ \sum_{i=1}^{n} \prod_{k=1}^{3} \Big(\sum_{|i_k - i| \le b} \check{e}_{i_kn}^2 \Big) \Big\}^{1/2} \Big] \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \\
&\le C(c_1,\delta) \sum_{b=0}^{n-1} b^{3/2} \Big[\Big\{ \sum_{i=1}^{n} \check{e}_{in}^2 \Big\}^{1/2} \Big\{ \Big(b \sum_{i=1}^{n} \check{e}_{in} \Big) \Big(b \max\{\check{e}_{in}^2 : 1 \le i \le n\} \Big)^2 \Big\} \Big] \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \\
&\le C(c_1,\delta) \Big[\sum_{b=0}^{n-1} b^3 \alpha(b;\theta)^{\frac{\delta}{4+\delta}} \Big] \times \Big[\sum_{n=1}^{\infty} n^3 \alpha(n;\theta)^{\frac{\delta}{4+\delta}} \Big] \times \Big[\max\{\check{e}_{in}^2 : 1 \le i \le n\} \Big)^2 \Big\} \Big] \\
&= O(1) \quad \text{uniformly in} \quad \theta \in \Theta.
\end{aligned}
$$

6.2 Proofs of the main results:

Proof of Theorem 2.1: It is enough to show that,

$$
|M_{1n}(\tilde{\theta}_n) - M_{1n}(\theta_0)| + \sum_{i=2}^{3} \Big(|M_{in}(\tilde{\theta}_n)| + |M_{in}(\theta_0)| \Big) = o_p(1). \tag{6.3}
$$

Note that by (C.2) and the condition $\tilde{\theta}_n \xrightarrow{p} \theta_0$, $|M_{1n}(\tilde{\theta}_n) - M_{1n}(\theta_0)| = o_p(1)$ if,

$$
\gamma_n(\tilde{\theta}_n)' \Gamma_n^{-1}(\tilde{\theta}_n) \gamma_n(\tilde{\theta}_n) - \gamma(\theta_0)' \Gamma_n^{-1}(\theta_0) \gamma_n(\theta_0) = o_p(1). \tag{6.4}
$$

It is easy to check that the absolute value of the right side of (6.3) is bounded above by

$$
\begin{aligned}
&|\gamma_n(\tilde{\theta}_n)' (\Gamma_n^{-1}(\tilde{\theta}_n) - \Gamma_n^{-1}(\theta_0)) \gamma_n(\tilde{\theta}_n)| \\
&+2\|\gamma_n(\tilde{\theta}_n) - \gamma_n(\theta_0)\| \|\Gamma_n^{-1}(\theta_0)\| \Big(\|\gamma_n(\tilde{\theta}_n)\| + \|\gamma_n(\theta_0)\| \Big) \\
&\equiv I_{1n} + I_{2n}, \text{ say.}
\end{aligned} \tag{6.5}
$$

By using the standard isometric isomorphism between $\ell^2(\mathbb{Z})$ and $L^2(0,2\pi)$ through the Fourier-Plancherel transform (cf. Bhatia (2003), Rudin (1987)), we have,

$$
\begin{aligned}
\|\Gamma_n^{-1}(\theta)\| &\leq C\|f_\theta^{-1}\|_\infty \\
\|\Gamma_n(\theta_1) - \Gamma_n(\theta_2)\| &\leq C\|f_{\theta_1} - f_{\theta_2}\|_\infty, \text{ for all } \theta_1, \theta_1 \in \Theta, n \geq 1.
\end{aligned} \tag{6.6}
$$

By (6.6) and conditions (C.2) and (C.3),

$$
\begin{aligned}
I_{1n} &= |\gamma_n(\tilde{\theta}_n)'\Gamma_n^{-1}(\theta_0)(\Gamma_n(\tilde{\theta}_n) - \Gamma_n(\theta_0))\Gamma_n^{-1}(\tilde{\theta}_n)\gamma_n(\tilde{\theta}_n)| \\
&\leq \|\gamma_n(\tilde{\theta}_n)\|^2\|\Gamma_n^{-1}(\theta_0)\|\|\Gamma_n(\tilde{\theta}_n)\|\|\Gamma_n(\tilde{\theta}_n) - \Gamma_n(\theta_0)\| \\
&= o_p(1).
\end{aligned}
$$

By similar arguments, on the set $\{\|\tilde{\theta}_n - \theta_0\| < \epsilon\},\ (0 < \epsilon < \delta)$,

$$
\begin{aligned}
I_{2n}^2 &\leq C\|\gamma_n(\tilde{\theta}_n) - \gamma_n(\theta_0)\|^2 \\
&= C\left[\sum_{j=0}^{M-1} |g_j(\tilde{\theta}_n) - g_j(\theta_0)|^2 + \sum_{j=M}^{n-1} |g_j(\tilde{\theta}_n) - g_j(\theta_0)|^2\right] \\
&\leq C\left[\sum_{j=0}^{M-1} \sup_{\|x\|\leq\epsilon} |g_j(\theta_0 + x) - g_j(\theta_0)| + \sum_{j=M}^{\infty} \sup_{\theta\in\Theta_0} |g_j(\theta)|\right]
\end{aligned}
$$

Given any $\eta > 0$, there exist $M \geq 2$, such that, $\sum_{j=M}^{\infty} \sup_{\theta\in\Theta_0} |g_j(\theta)| < \frac{\eta}{[3C]}$. Next, given $M \geq 1$ and $\eta > 0$, there exists $\epsilon \in (0, \delta)$ such that

$$
\sup_{\|x\|\leq\epsilon} |g_j(\theta_0 + x) - g_j(\theta_0)| < \frac{\eta}{3MC}, \text{ for all } j = 0, \ldots, M.
$$

Hence,

$$
\begin{aligned}
P(I_{2n}^2 > \eta) &\leq P(\|\tilde{\theta}_n - \theta_0\| > \epsilon) + P(I_{2n}^2 > \eta, \|\tilde{\theta}_n - \theta_0\| < \delta) \\
&\leq P(\|\tilde{\theta}_n - \theta_0\| > \epsilon) + 0 \quad \text{for large} \quad n \\
&= o(1).
\end{aligned}
$$

By similar arguments,

$$
\begin{aligned}
P(M_{3n}(\tilde{\theta}_n) > \epsilon) &\leq P(\sup\{M_{3n}(\theta) : \theta \in \Theta_0\} > \epsilon, \tilde{\theta}_n \in \Theta_0 + P(\tilde{\theta}_n \notin \Theta_0) \\
&= o(1).
\end{aligned}
$$

Since $M_{2n}(\theta) \leq 2\Big[M_{1n}(\theta)M_{3n}(\theta)\Big]^{1/2}$ for all θ, the theorem is proved.

Proof of (3.6)**:** Note that by Lemma 6.3, $\sup\{|\tilde{\mu}_{3n}(\theta)| : \theta \in \Theta\} = O(1)$. Hence, noting that $\sup\{(E_\theta\|R_n\|^8)^{1/8} : \theta \in \Theta\} = O(d_n) = o(n^{-1/2})$, it is enough to show that

$$\sup\left\{\left|\tilde{\mu}_{3n}(\theta) - E_\theta\left(\Big[n^{-1/2}\beta_0(\theta) + n^{-1/2}\sum_{i=1}^{n}\xi_i\Big]'\boldsymbol{\Delta\lambda}_n(\theta)\mathbf{X}_n\right)^2\right| : \theta \in \Theta\right\} = o(1). \qquad (6.7)$$

Now expanding the second term and applying the first part of Lemma 6.3, one can conclude that the left side of (6.7) is in fact $O(n^{-1})$. This completes the proof of (3.6).

Proof of Theorem 3.1: By (C.1), there exists $C \in (0,\infty)$ such that $\sup\{|M_{1n}^{(1)}(\theta)|^{-1} : \theta \in \Theta_0, j,l = 1,\cdots,p; n \geq 1\} < C$. Let

$$\hat{D}_n = \sum_{j=1}^{p} M_{1n}^{(j)}(\hat{\theta}_n)\hat{\beta}_{n,j} + \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l)M_{1i}^{(j,l)}(\hat{\theta}_n)\hat{\Sigma}_n(j,l)$$
$$\tilde{D}_n = \sum_{j=1}^{p} M_{1i}^{(j)}(\theta_0)\hat{\beta}_{n,j} + \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l)M_{1i}^{(j,l)}(\theta_0)\hat{\Sigma}_n(j,l),$$

$n \geq 1$, where $w(j,l) = 1/2$ for $j \neq l$ and $w(j,l) = 1$ for $j = l$. Then by Taylor's expansion, it follows that there exists a constant $C \in (0,\infty)$ such that on the set $\{\hat{\theta} \in \Theta_0\}$,

$$\hat{D}_n = \tilde{D}_n + R_{1n}, \quad \text{and} \quad \mathbf{r}_n = -\frac{\tilde{D}_n}{M_{1n}^{(1)}(\theta_0)}\boldsymbol{e}_1 + R_{2n}\boldsymbol{e}_1 \qquad (6.8)$$

where $|R_{1n}| \leq C\Big\{\|\hat{\beta}_n\|.\|\hat{\theta}_n - \theta_0\| + \|\hat{\theta}_n - \theta_0\|^\gamma\|\hat{\Sigma}_n\|\Big\}$ and $|R_{2n}| \leq C\Big\{|\hat{D}_n|.\|\hat{\theta}_n - \theta_0\| + |R_{1n}|\Big\}$.

Let $A_{1n} \equiv \{\hat{\theta}_n \in \Theta_0\} \cap \{\hat{\theta}_n + \mathbf{r}_n \in \Theta\}$. Using similar arguments, on the set A_{1n}, for all $u \in [0,1]$, we have

$$\sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l)M_{1n}^{(jl)}(\hat{\theta}_n + u\mathbf{r}_n)\Big([\hat{\theta}_n + \mathbf{r}_n] - \theta_0\Big)^{\boldsymbol{e}_j + \boldsymbol{e}_l}$$
$$= \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l)M_{1n}^{(jl)}(\theta_0)\Big(\hat{\theta}_n - \theta_0\Big)^{\boldsymbol{e}_j + \boldsymbol{e}_l} + R_{3n}(u)$$

where $\sup_{u\in[0,1]}|R_{3n}(u)| \leq C\Big[\|(\hat{\theta}_n + \mathbf{r}_n) - \theta_0\|^{2+\gamma} + \|(\hat{\theta}_n + \mathbf{r}_n) - \theta_0\| \cdot \|\hat{\theta}_n - \theta_0\| + \|\mathbf{r}_n\|^2\Big]$ for some $C \in (0,\infty)$.

Next define the set $A_{2n} = A_{1n} \cap \{\hat{\theta}_n + \mathbf{r}_n \in \Theta_0\}$. Then, on $A_{2n} = \{\hat{\theta}_n, \hat{\theta}_n + \mathbf{r}_n \in \Theta_0\}$, by

Taylor's expansion, there exists a point $\theta_n^\dagger$ on the line joining $\hat{\theta}_n + \mathbf{r}_n$ and θ_0 such that

$$
\begin{aligned}
& M_{1i}(\hat{\theta}_n + \mathbf{r}_n) - M_{1i}(\theta_0) \\
&= \sum_{j=1}^{p} M_{1n}^{(j)}(\theta_0)\Big\{[\hat{\theta}_n + \mathbf{r}_n] - \theta_0\Big\}^{\boldsymbol{e}_j} + \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l) M_{1i}^{(jl)}(\theta_n^\dagger)\Big\{[\hat{\theta}_n + \mathbf{r}_n] - \theta_0\Big\}^{\boldsymbol{e}_j+\boldsymbol{e}_l} \\
&= \sum_{j=1}^{p} M_{1n}^{(j)}(\theta_0)\Big(\hat{\theta}_n - \theta_0\Big)^{\boldsymbol{e}_j} + M_{1n}^{(1)}(\theta_0)\left(-\frac{\tilde{D}_n}{M_{1n}^{(1)}(\theta_0)} + R_{2n}\right) \\
&\qquad + \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l) M_{1n}^{(jl)}(\theta_0)\Big\{\hat{\theta}_n - \theta_0\Big\}^{\boldsymbol{e}_j+\boldsymbol{e}_l} + R_{3n}^\dagger \\
&= \sum_{j=1}^{p} M_{1n}^{(j)}(\theta_0)\left\{\Big(\hat{\theta}_n - \theta_0\Big)^{\boldsymbol{e}_j} - \hat{\beta}_{n,j}\right\} \\
&\qquad + \sum_{j=1}^{p}\sum_{l=1}^{p} w(j,l) M_{1i}^{(j,l)}(\theta_0)\left\{\Big(\hat{\theta}_n - \theta_0\Big)^{\boldsymbol{e}_j+\boldsymbol{e}_l} - \hat{\Sigma}_n(j,l)\right\} + M_{1n}^{(1)}(\theta_0)R_{2n} + R_{3n}^\dagger \\
&\equiv Q_{1n} + M_{1n}^{(1)}(\theta_0)R_{2n} + R_{3n}^\dagger, \text{ say}
\end{aligned}
\tag{6.9}
$$

where $R_{3n}^\dagger = R_{3n}(u)$ with the u corresponding to $\theta_n^\dagger$.

Hence, on the set $A_{3n} \equiv \{|M_{1n}^{(1)}(\hat{\theta}_n)|^{-1} \le (1+\log n)^2\}$,

$$
\begin{aligned}
& M_{1n}(\check{\theta}_n) - M_{1n}(\theta_0) \\
&= [M_{1n}(\hat{\theta}_n + \mathbf{r}_n) - M_{1n}(\theta_0)]\mathbb{1}\Big(\Big\{\hat{\theta}_n + \mathbf{r}_n \in \Theta\Big\} \cap A_{3n}\Big) \\
&\qquad + [M_{1n}(\hat{\theta}_n) - M_{1n}(\theta_0)]\mathbb{1}\Big(\Big\{\hat{\theta}_n + \mathbf{r}_n \notin \Theta\Big\} \cup A_{3n}^c\Big) \\
&= [M_{1n}(\hat{\theta}_n + \mathbf{r}_n) - M_{1n}(\theta_0)]\Big\{\mathbb{1}(A_{2n}) + \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta) - \mathbb{1}(A_{2n})\Big\}\mathbb{1}(A_{3n}) \\
&\qquad + [M_{1n}(\hat{\theta}_n) - M_{1n}(\theta_0)]\mathbb{1}\Big(\Big\{\hat{\theta}_n + \mathbf{r}_n \notin \Theta\Big\} \cup A_{3n}^c\Big) \\
&\equiv \Big[Q_{1n} + M_{1n}^{(1)}(\theta_0)R_{2n} + R_{3n}^\dagger\Big]\mathbb{1}(A_{2n} \cap A_{3n}) + R_{4n}, \text{ say} \\
&\equiv Q_{1i} + R_{5n}, \text{say},
\end{aligned}
\tag{6.10}
$$

where $|R_{5n}| \le |R_{4n}| + |R_{2n} + R_{3n}^\dagger|\mathbb{1}(A_{2n}) + |Q_{1n}|\mathbb{1}(A_{2n}^c \cap A_{3n}^c)$ and

$$
\begin{aligned}
|R_{4n}| &\le \Big|M_{1n}(\hat{\theta}_n + \mathbf{r}_n) - M(\theta_0)\Big| \cdot \Big|\mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta) - \mathbb{1}(A_{2n})\Big|\mathbb{1}(A_{3n}) \\
&\qquad + \Big|M_{1n}(\hat{\theta}_n - M_{1n}(\theta_0)\Big|\mathbb{1}(\{\hat{\theta}_n + \mathbf{r}_n \notin \Theta_0\} \cup A_{3n}^c) \\
&\equiv R_{41n}, \text{ say.}
\end{aligned}
$$

Note that by definition,

$$\begin{aligned}
&\Big|\mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta) - \mathbb{1}(A_{2n})\Big| \\
&\le \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta)\mathbb{1}(A_{2n}^c) + \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \notin \Theta)\mathbb{1}(A_{2n}) \\
&\le \{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta \setminus \Theta_0)\} + \mathbb{1}(\emptyset).
\end{aligned}$$

Hence, with $A_{4n}^c \equiv \{\hat{\theta}_n + \mathbf{r}_n \notin \Theta_0\} \cap A_{3n}$,

$$\begin{aligned}
R_{41n} \le\ & \Big|M_{1n}(\hat{\theta}_n + \mathbf{r}_n) - M_{1n}(\hat{\theta}_n)\Big|\mathbb{1}(A_{3n})\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta \setminus \Theta_0)\} \\
& +2\Big|M_{1n}(\hat{\theta}_n) - M_{1n}(\theta_0)\Big|\Big\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}\Big(\{\hat{\theta}_n + \mathbf{r}_n \notin \Theta_0\} \cap A_{3n}\Big) + \mathbb{1}(A_{3n}^c)\Big\} \\
\le\ & C\|\mathbf{r}_n\|\mathbb{1}(A_{3n})\Big\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(\hat{\theta}_n + \mathbf{r}_n \in \Theta \setminus \Theta_0)\Big\} \\
& +C\|\hat{\theta}_n - \theta_0\|\Big\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(A_{4n}^c) + \mathbb{1}(A_{3n}^c)\Big\} \\
\le\ & C \cdot (\log n)^2\{\|\hat{\beta}_n\| + \|\hat{\Sigma}_n\|\}\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(A_{4n}^c)\} \\
& +C \cdot \|\hat{\theta}_n - \theta_0\|\Big\{\mathbb{1}(\hat{\theta}_n \notin \Theta_0) + \mathbb{1}(A_{4n}^c) + \mathbb{1}(A_{3n}^c)\Big\}. \qquad (6.11)
\end{aligned}$$

By condition ??, there exist $C \in (0, \infty)$ and $\epsilon_1 \in (0, \frac{\epsilon_0}{2})$ such that

$$\begin{aligned}
A_{4n}^c &\subset \{\|\hat{\theta}_n - \theta_0\| > \frac{\epsilon_0}{2}\} \cup \{\|\mathbf{r}_n\| > \frac{\epsilon_0}{2}\} \\
&\subset \{\|\hat{\theta}_n - \theta_0\| > \epsilon_1\} \cup \{(\log n)^2(\|\hat{\beta}_n\| + \|\hat{\Sigma}_n\|) > C\} \qquad (6.12)
\end{aligned}$$

and $A_{3n}^c \subset \{\|\hat{\theta}_n - \theta_0\| > \epsilon_1\}$ for all $n \ge 1$. Hence, it follows that

$$\begin{aligned}
R_{41n} \le\ & C \cdot (\log n)^2\{\|\hat{\beta}_n\| + \|\hat{\Sigma}_n\|\}\Big[\mathbb{1}(\|\hat{\theta}_n - \theta_0\| > \epsilon_1) + \mathbb{1}\Big([\log n]^2(\|\hat{\beta}_n\| + \|\hat{\Sigma}_n\|) > C\Big)\Big] \\
& +C \cdot \|\hat{\theta}_n - \theta_0\|\Big[\mathbb{1}(\|\hat{\theta}_n - \theta_0\| > \epsilon_1) + \mathbb{1}\Big([\log n]^2(\|\hat{\beta}_n\| + \|\hat{\Sigma}_n\|) > C\Big)\Big] \qquad (6.13)
\end{aligned}$$

for all $n \ge 1$. Let $W_n = (n\|\hat{\beta}_n\| + n\|\hat{\Sigma}_n\|)$. Note that by uniform integrability of $\{(\sqrt{n}\|\hat{\theta} - \theta\|)^2\}_{m\ge 1}$ and the fact that $E \mid W_n \mid^{1+\eta} = O(1)$,

$$\begin{aligned}
& E(R_{41n}) \\
\le\ & Cn^{-1}(\log n)^2\Big[\Big(E \mid W_n \mid^{1+\eta}\Big)^{\frac{1}{1+\eta}}\Big(P(\|\hat{\theta}_n - \theta_0\| > \epsilon_1\Big)^{\frac{\eta}{1+\eta}} \\
& \qquad +E \mid |W_n \mid^{1+\eta} \{n^{-1}(\log n)^2\}^{\eta}\Big] \\
& +C\Big[\epsilon_1^{-1}E\|\hat{\theta}_n - \theta_0\|^2\mathbb{1}(\|\hat{\theta}_n - \theta_0\| > \epsilon_1) \\
& \qquad +\Big(E\|\hat{\theta}_n - \theta_0\|^2\Big)^{1/2}\Big\{P(n^{-1}(\log n)^2|W_n| > C)\Big\}^{\frac{1}{2}}\Big] \\
=\ & o(n^{-1}) \quad \text{as} \quad m \to \infty. \qquad (6.14)
\end{aligned}$$

This proves the first part of Theorem 3.1.

Next we consider the bound on the variance of the tilted MSPE estimator. Since $\sup\{|M_{kn}(\theta)|^2 : \theta \in \Theta\} = O(n^{-2})$ for $k = 2, 3$, by Cauchy-Schwarz inequality, it is enough to show that

$$\mathrm{Var}\Big(M_{1n}(\check{\theta}_n)\Big) = O(n^{-1}). \tag{6.15}$$

By Taylor's expansion,

$$M_{1n}(\check{\theta}_n) = M_{1n}(\theta_0) + \sum_{j=1}^{p} M_{1n}^{(j)}(\theta_0)[\check{\theta}_n - \theta_0]^{\boldsymbol{e}_j} + R_{6n}$$

where $|R_{6n}| \le C(p)\Delta\|\check{\theta}_n - \theta_0\|^2$ and $\Delta_r = \limsup_{n\to\infty} \sup\{|M_{1n}^{\alpha}(\theta)| : \theta \in \Theta, |\alpha| = r\}, r = 1, 2$. Also, let $A_{5n} = \{\hat{\theta}_n + \mathbf{r}_n \in \Theta, |M_{1n}^{(1)}(\hat{\theta}_n)|^{-1} \le (1 + \log n)^2\}$. Thus, it follows that

$$\begin{aligned}
ER_{6n}^2 &\le C(p, \Delta_2)E\|\check{\theta}_n - \theta_0\|^4 \\
&= C(p, \Delta_2)\Big[E\|\hat{\theta}_n + \mathbf{r}_n - \theta_0\|^4 \mathbb{1}(A_{5n}) + E\|\hat{\theta}_n - \theta_0\|^4 \mathbb{1}(A_{5n}^c)\Big] \\
&\le C(p, \Delta_2)2^3\Big[E\|\hat{\theta}_n - \theta_0\|^4 + E\|\mathbf{r}_n\|^4 \mathbb{1}(A_{5n})\Big] \\
&\le C(p, \Delta_0, \Delta_1, \Delta_2)\Big[E\|\hat{\theta}_n - \theta_0\|^4 + (1 = \log n)^8 n^{-4}\Big] \\
&= O(n^{-2}).
\end{aligned} \tag{6.16}$$

By similar arguments and Cauchy-Schwarz inequality,

$$\begin{aligned}
&E\Big[\check{\theta}_n - \theta_0\Big]^{\boldsymbol{e}_i + \boldsymbol{e}_j} \\
&= E\Big[\hat{\theta}_n - \theta_0\Big]^{\boldsymbol{e}_i + \boldsymbol{e}_j} + O\Big(E\|\mathbf{r}_n\|^2 \mathbb{1}(A_{5n}) + \Big\{E\|hthn - \theta_0\|^2\Big\}^{1/2}\Big\{E\|\mathbf{r}_n\|^2 \mathbb{1}(A_{5n})\Big\}^{1/2} \\
&= O(n^{-1}) + O(n^{-3/2}[\log n]^2.
\end{aligned}$$

Hence, it follows that

$$\begin{aligned}
&\mathrm{Var}\Big(\sum_{j=1}^{p} M_{1n}^{(j)}(\theta_0)[\check{\theta}_n - \theta_0]^{\boldsymbol{e}_j}\Big) \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} M_{1n}^{(i)}(\theta_0)M_{1n}^{(j)}(\theta_0)\mathrm{Cov}\Big([\check{\theta}_n - \theta_0]^{\boldsymbol{e}_i}, [\check{\theta}_n - \theta_0]^{\boldsymbol{e}_j}\Big) \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} M_{1n}^{(i)}(\theta_0)M_{1n}^{(j)}(\theta_0)\mathrm{Cov}\Big([\check{\theta}_n - \theta_0]^{\boldsymbol{e}_i}, [\check{\theta}_n - \theta_0]^{\boldsymbol{e}_j}\Big) \\
&= O(n^{-1}).
\end{aligned} \tag{6.17}$$

Hence, by (6.16), (6.17), and Cauchy-Schwarz inequality, (6.15) follows. This completes the proof of Theorem 3.1.

Proof of Proposition 4.1: For $b = 1, \ldots, B$, let $\Upsilon_{1j}^{*b} = \tilde{\Psi}_n^{*b}(\hat{\theta}_n + a_n \boldsymbol{e}_j) - \Psi_n^{*b}(\hat{\theta}_n + a_n \boldsymbol{e}_j)$ and let Υ_{2j}^{*b} be defined by replacing $\hat{\theta}_n + a_n \boldsymbol{e}_j$ by $\hat{\theta}_n - a_n \boldsymbol{e}_j$ in Υ_{1j}^{*b}, $1 \le j \le p$. Then, by Taylor's expansion

$$\begin{aligned}
&\left| E_* M_{1n}^{*(j)}(\hat{\theta}_n) - M_{1n}(\hat{\theta}_n) \right| \\
&= \left| (2a_n)^{-1} \left[M_{1n}(\hat{\theta}_n + a_n \boldsymbol{e}_j) - M_{1n}(\hat{\theta}_n - a_n \boldsymbol{e}_j) \right] - M_{1n}(\hat{\theta}_n) \right| \\
&\le C a_n \sup\{ M_{1n}(\theta) : \theta \in \Theta \}.
\end{aligned}$$

Next, by (conditional) independence of $\{\Upsilon_{kj}^{*b} : b = 1, \ldots, B\}$, $k = 1, 2$,

$$\text{Var}_*([2Ba_n]^{-1} \sum_{b=1}^{B} [\Upsilon_{1j}^{*b} - \Upsilon_{2j}^{*b}]) = O(a_n^{-2} B^{-1}), \quad k = 1, 2.$$

This proves the first part of Proposition 4.1. The proof of the second part is similar and hence, is omitted.

Proof of Proposition 4.2: Similar to the proof of Proposition 4.1 and hence is omitted.

Acknowledgement The authors thank as anonymous referee for some constructive suggestions that improved the exposition of the paper.

References

Ansley, C. F. and Kohn, R. (1986). Prediction Mean Squared Error for State Space Models with Estimated Parameters. *Biometrika* **73** 467–473.

Bhatia, R. (2003). *Fourier Series.* Hindusthan Book Agency, New delhi, India.

Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*, Second Edition. Springer, New York.

Lahiri, S. N., (2003). *Resampling Methods for Dependent Data*, Springer, New York.

Lahiri, S. N. and Maiti, T. (2003). Nonnegative Mean Squared Error prediction. Preprint. Posted at http://arxiv.org/abs/math.ST/0604075.

Lahiri, S. N., Maiti, T., Katzoff, M., and Parsons, V. (2007). Resampling based Empirical Prediction: An Application to Small Area Estimation. *Biometrika* **94** 469–485.

Pfeffermann, D. and Tiller, R. (2005). Bootstrap Approximation to Prediction MSE for State-Space Models with Estimated Parameters. *Journal of Time Series Analysis* **26** 893–916.

Prasad, N. G. N. and Rao, J. N. K. (1990). The Estimation of the Mean Squared Error of Small-Area Estimators. *Journal of the American Statistical Association* **85** 163–171.

Priestley, M. B. (1981). *Spectral analysis and time series*, Vols. 1 and 2. Academic Press, New York.

Quenneville, B. and Singh, A. C. (2000). Bayesian Prediction Mean Squared Error for State Space Models with Estimated Parameters. *Journal of Time Series Analysis* **21** 219 – 236.

Rao, J. N. K. (2003). *Small Area Estimation*, Wiley, New York.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill Book Co., New york.

		Not bias-corrected		Bias-corrected	
n	Model	Mean	RMSE	Mean	RMSE
50	1	0.966	0.976	0.547	1.216
	2	0.534	0.940	0.488	0.964
120	1	0.317	0.564	0.103	0.568
	2	0.169	0.547	0.135	0.604
500	1	0.043	0.297	0.003	0.353
	2	0.098	0.311	0.049	0.377

Table 1: *Bias and Root mean squared error (RMSE) for the estimators (with and without bootstrap based bias correction) of the mean squared prediction errors for models in (5.1)-(5.2) for sample size(n) = 50, 120 and 500, number of replications(N) = 500 and number of bootstrap samples(N_0) = 1000.*

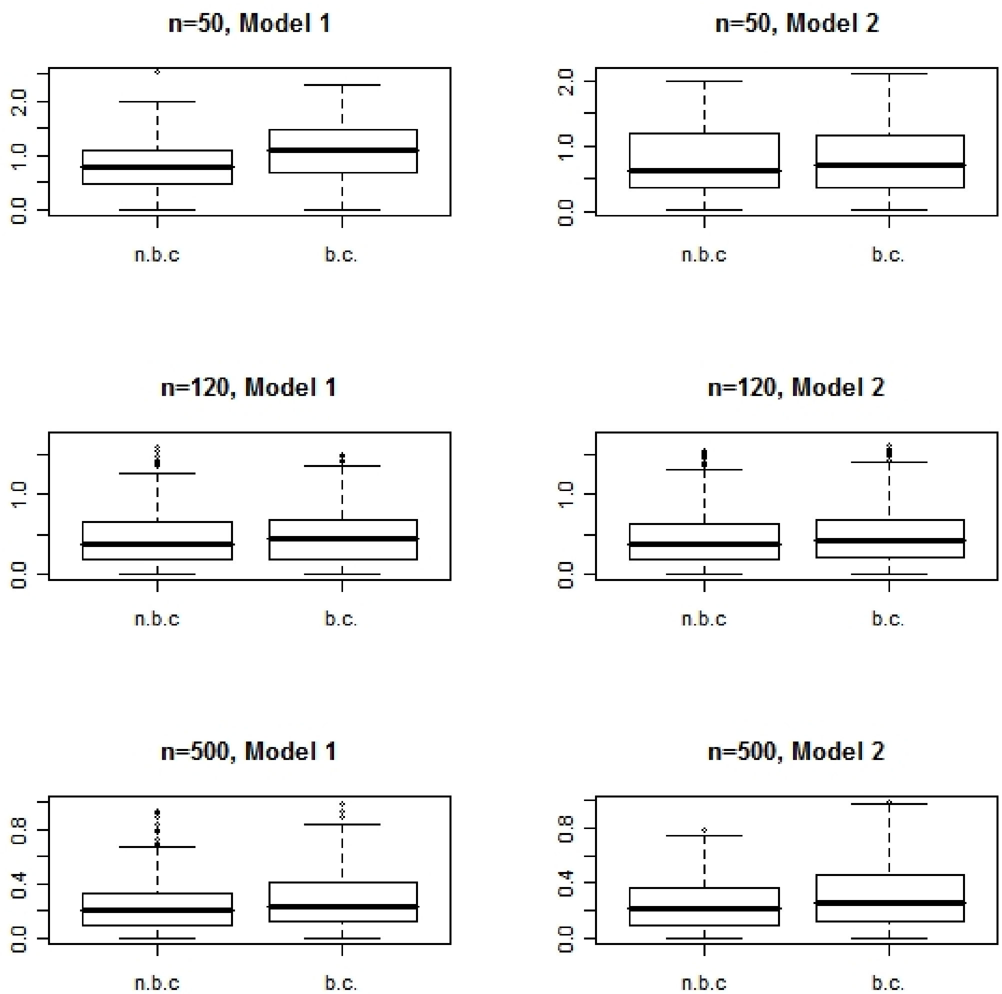

Figure 1: *Boxplots of RMSE values of the estimators (with bias-correction (b.c.) and without bias correction (n.b.c.)) of the mean squared prediction errors for n=50, 120 and 500 under the three models as in (5.1)-(5.2)*