Bootstrap of minimum distance estimators in regression with correlated disturbances

Juan M. Vilar-Fernández *, José A. Vilar-Fernández

Departamento de Matemáticas, Facultad de Informática, Universidad de A Coruña, Campus de Elviña, 15071 A Coruña, Spain

Abstract

The linear regression model \( Y_i = x_i \theta + \varepsilon_i \), where \( x_i \in \mathbb{R}^p \), \( \theta \) is an unknown parameter vector and the observational errors \( \varepsilon_i \) follow an AR(1) model, is considered. In a previous paper, Vilar-Fernández and González-Manteiga (Statist. Papers, to appear) have proposed a two-stage generalized minimum distance estimator (GMD) for \( \theta \), which presents the same asymptotic properties as the generalized least-squares estimator (GLS). However, for finite samples, the GMD estimator has been shown to be more efficient in terms of the mean-squared error than the GLS one when the errors of the model are heavily correlated and with large variance.

This paper establishes the consistency of a bootstrap procedure to approximate the distribution of the GMD estimator. In addition, the good behavior of the bootstrap approximation with respect to the asymptotic distribution of the GMD is shown in a simulation study.

© 2002 Elsevier Science B.V. All rights reserved.

MSC: primary 62J05; secondary 62G05

Keywords: Minimum distance estimators; Linear regression models; Nonparametric estimation; Bootstrap

1. Introduction

Consider the following linear regression model:

\[
Y = X\theta + \varepsilon, \quad \theta \in \Theta \subset \mathbb{R}^p,
\]

where \( X = \{x_i\}_{i=1}^n = \{x_{ij}\} \), \( i = 1, \ldots, n \), \( j = 1, \ldots, p \), is the \( n \times p \) design matrix, with \( x_i \in C \), being \( C \) a compact set in \( \mathbb{R}^p \), \( Y = \{y_i\}_{i=1}^n \) is the response variable vector of dimension \( n \times 1 \) and \( \varepsilon = \{\varepsilon_i\}_{i=1}^n \) is the random error vector, which we assume
follows an AR(1)-type correlation structure, \( e_i = \rho e_{i-1} + \varepsilon_i, \ i \in \mathbb{Z} \), with \( |\rho| < 1 \) and \( \{\varepsilon_i\}_{i \in \mathbb{Z}} \) a continuously distributed noise process, with zero mean, finite variance \( \sigma^2 \) and distribution function \( F \).

These models frequently arise in economic studies, in the analysis of growth curves, and in situations in which data are sequentially collected over time (see Amemiya, 1989; Judge et al., 1985, as good references among many others). In order to estimate the parameter vector \( \theta \), Vilar-Fernández and González-Manteiga (2000) have proposed a method that combines the ideas of generalized least-squares (GLS) estimation (see Judge et al., 1985; Chapter 8), transforming the statistical model into another one with uncorrelated errors, and minimum distance estimation, based on constructing a nonparametric pilot estimator of the regression function. More precisely, the new method consists of the following stages:

I. Compute a consistent estimator of \( \theta \), \( \hat{\theta}_{\text{MD}} \), using the minimum distance method. That is, \( \hat{\theta}_{\text{MD}} \) is defined as the vector that minimizes a Crámer–von-Mises-type functional distance between the regression function, \( m(x) = E(Y/X = x) = x\theta \), and a nonparametric estimator of \( m \), \( \hat{m}_{n,h}(x) \), having the general form \( \hat{m}_{n,h}(x) = \sum_{i=1}^{n} \omega_{ni,h}(x_i)Y_i \), where the weight sequence \( \{\omega_{ni,h}(x_i)\}_{i=1}^{n} \) depends on the smoothing parameter \( h = h(n) \).

So, the \( \hat{\theta}_{\text{MD}} \) estimator is defined as the argument that minimizes the functional

\[
\Psi(\theta) = \int (\hat{m}_{n,h}(x) - x^t\theta)^2 d\Omega_n(x) = \frac{1}{n} \sum_{i=1}^{n} (\hat{m}_{n,h}(x_i) - x_i^t\theta)^2,
\]

(2)

where \( \Omega_n(x) \) is the empirical distribution of the design points \( \{x_i\}_{i=1}^{n} \).

Thus, the estimator \( \hat{\theta}_{\text{MD}} \) is given by

\[
\hat{\theta}_{\text{MD}} = (X^tX)^{-1}X^tM_h = \left( \sum_{i=1}^{n} x_i x_i^t \right)^{-1} \left( \sum_{i=1}^{n} x_i \hat{m}_{n,h}(x_i) \right),
\]

(3)

where

\[
M_h = \begin{pmatrix}
\hat{m}_{n,h}(x_1) \\
\vdots \\
\hat{m}_{n,h}(x_n)
\end{pmatrix} = \begin{pmatrix}
\sum_{i=1}^{n} \omega_{n1,h}(x_1)Y_i \\
\vdots \\
\sum_{i=1}^{n} \omega_{nn,h}(x_n)Y_i
\end{pmatrix}
\]

\[
= \begin{pmatrix}
\omega_{n1,h}(x_1) & \cdots & \omega_{nn,h}(x_n)
\end{pmatrix}
\begin{pmatrix}
Y_1 \\
\vdots \\
Y_n
\end{pmatrix} = W_h Y,
\]

so that \( W_h \) denotes the smoothing matrix associated with the nonparametric estimator \( \hat{m}_{n,h} \).

The estimator \( \hat{\theta}_{\text{MD}} \) has been studied by Cristóbal et al. (1987), Stute and González-Manteiga (1990), González-Manteiga (1995) and Akritas (1996), among others, in a context of independent observations, and by González-Manteiga and Vilar-Fernández (1995) with correlated errors.
II. The residuals are obtained as follows:

$$\tilde{e}_i = Y_i - x'_i \hat{\theta}_{\text{MD}}, \quad i = 1, 2, \ldots, n,$$

and then a consistent estimator for $\rho$ is computed from the above residuals

$$\hat{\rho} = \frac{\sum_{i=1}^{n-1} \tilde{e}_i \tilde{e}_{i+1}}{\sum_{i=1}^{n} \tilde{e}_i^2}.$$  \hspace{1cm} (4)

III. The observations are transformed to obtain a model with uncorrelated errors. In our case, the transformation matrix, $P$, is

$$P = \begin{pmatrix}
\sqrt{1 - \rho^2} & 0 & 0 & \ldots & 0 \\
-\rho & 1 & 0 & \ldots & 0 \\
0 & -\rho & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & -\rho & 1
\end{pmatrix}.$$  

For this matrix $P$, we have

$$P'P = A^{-1} = \begin{pmatrix}
1 & -\rho & 0 & 0 & \ldots & 0 \\
-\rho & 1 + \rho^2 & -\rho & 0 & \ldots & 0 \\
0 & -\rho & 1 + \rho^2 & -\rho & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & -\rho & 1
\end{pmatrix},$$  

where $A$ is the correlation matrix of the errors $\{e_i\}_{i=1}^n$. Since $P$ and $A^{-1}$ only depend on $\rho$, natural estimators for these matrices ($\hat{P}$ and $\hat{A}^{-1}$) can be obtained by replacing $\rho$ by its estimator $\hat{\rho}$. The transformed model is

$$\hat{Y} = \hat{P}Y = \hat{P}X\theta + \hat{P}e = \hat{X}\theta + e.$$  \hspace{1cm} (5)

IV. The minimum distance method is applied to the transformed observations $(\hat{X}, \hat{Y})$, with $\hat{X} = \hat{P}X$ and $\hat{Y} = \hat{P}Y$, to obtain the generalized minimum distance (GMD) estimator, $\hat{\theta}_{\text{GMD}}$. Therefore, $\hat{\theta}_{\text{GMD}}$ has the following form

$$\hat{\theta}_{\text{GMD}} = (X'\hat{A}^{-1}X)^{-1}(X'\hat{P}'W_g\hat{P}Y),$$  \hspace{1cm} (6)

where $W_g = \{\omega_{m,g}(\hat{x}_j)\}_{j=1}^{n}$ is the new smoothing matrix associated with the nonparametric estimator of $m$, which has now been constructed from the transformed data and the smoothing parameter $g = g(n)$.

The strong consistency and the asymptotic distribution of $\hat{\theta}_{\text{GMD}}$ were established in Vilar-Fernández and González-Manteiga (2000) under appropriate general assumptions. More precisely, in their Theorem 6 it is stated that $n^{1/2}(\hat{\theta}_{\text{GMD}} - \theta) \rightarrow N(0, \Sigma)$ in distribution, where $\Sigma$ denotes the limit covariance matrix. Such a result, with the same asymptotic covariance matrix, has been already obtained for the GLS estimator (which from now on will be denoted by $\hat{\theta}_{\text{GLS}}$) in Lemma 3.1 in Stute (1995) or in an earlier paper of Amemiya (1973). However, although both estimators present the same limit distribution, the simulation study performed in Vilar-Fernández and González-Manteiga (2000) allows us to observe that, for finite samples, the $\hat{\theta}_{\text{GMD}}$ estimator exhibits a better
behavior than the $\hat{\theta}_{\text{GLS}}$ estimator in terms of mean-squared error, when suitable choices of the smoothing parameters are used in the nonparametric estimation of $m(x)$.

In general, the minimum distance estimators are likely to outperform least-squares estimators when the variance of the error is large, that is, the signal-to-noise ratio is small. In other cases, the estimation methods are almost equivalent since the effect of smoothing is negligible (Stute and González-Manteiga, 1990).

One of the drawbacks in the computation of the $\hat{\theta}_{\text{GMD}}$ estimator is that it depends on two smoothing parameters: a first parameter $h$ is necessary to obtain $\hat{\theta}_{\text{MD}}$ and then a second parameter $g$ must be chosen in order to apply again the minimum distance method to the transformed data. The parameter $h$ has a small influence on the calculation of $\hat{\theta}_{\text{GMD}}$ (see Vilar-Fernández and González-Manteiga, 2000). In practice, it is reasonable to take as pilot parameter $h = 0$ to simplify the computation of the estimator, that is, to use the least-squares residuals to estimate $\rho$ in the second stage. In fact, the aim of stages I and II is to obtain a consistent estimate for $\rho$, which can be made by means of different procedures to those presented in this paper (see, e.g. Herrmann et al., 1992; Judge et al., 1985, Chapter 8). However, the choice of the second parameter, $g$, is very important since it has a significant influence on the result of the estimation. Vilar-Fernández and González-Manteiga (2000) have proposed a cross-validation type method to compute $g$ from the data.

As in generalized least-squares estimation the proposed estimator could be improved using an iterative algorithm. For this purpose, stages II–IV might be successively repeated so that the residuals $\tilde{e}_i$ in stage II would be updated from the estimator $\hat{\theta}_{\text{GMD}}$ calculated in the last iteration.

In this work, we propose a bootstrap procedure to approximate the sampling distribution of $n^{1/2}(\hat{\theta}_{\text{GMD}} - \theta)$ and the main result of this paper shows the consistency in distribution of the bootstrap with probability one. Similar problems have been studied by a great number of authors: Freedman (1981) considered the bootstrap of a linear model with i.i.d. errors, Freedman (1984) and Stute (1995) developed a bootstrap approximation of the distribution of the two-stage GLS estimator, Kreiss and Franke (1992) studied the bootstrap of ARMA models and Datta and McCormick (1995) and Vilar-Fernández and González-Manteiga (1996) considered the bootstrap of a linear model with dependent errors to approximate the distribution of the minimum distance estimator.

An outline of the present paper is as follows: in Section 2, the resampling method is described and several simulation studies are presented to show the good behavior of the bootstrap. In Section 3, the main results that prove the validity and consistency of the bootstrap are obtained. The last section is devoted to the proof of the main result.

2. The bootstrap procedure

We now describe the resampling mechanism, which consists of the following steps: 

Step 1: Given the initial sample $\{(x_i, Y_i)\}_{i=1}^n$, the $\hat{\theta}_{\text{GMD}}$ estimator given in (6) is computed by following the stages described in Section 1.
Step 2: The new residuals $\hat{\varepsilon} = Y - X \hat{\theta}_{\text{GMD}}$ are evaluated and then the noise of the AR(1) model is obtained as $\tilde{\varepsilon} = \hat{P} \tilde{\varepsilon}$.

Step 3: The error series $\tilde{\varepsilon}$ is centered,

$$\tilde{\varepsilon}_i = \tilde{\varepsilon}_i - \tilde{\varepsilon}, \quad 2 \leq i \leq n,$$

where $\tilde{\varepsilon} = 1/(n-1) \sum_{i=2}^{n} \tilde{\varepsilon}_i$.

Step 4: The empirical distribution, $\hat{F}_n$, based on $\tilde{\varepsilon}_i$, $2 \leq i \leq n$, is derived.

Step 5: For a large enough number $M$, a sample of i.i.d. random variables, $\varepsilon^*_i$, with $-M \leq i \leq n$, is drawn from $\hat{F}_n$. Then, the values $\varepsilon^*_i$, for $1 \leq i \leq n$, are obtained by

$$\varepsilon^*_i = \sum_{j=0}^{\infty} \hat{\rho}^j \varepsilon^*_{i-j}.$$

In practice, an initial value $\varepsilon^*_0 = \sum_{j=0}^{M} \hat{\rho}^j \varepsilon^*_{i-j}$ is computed and then the equation $\varepsilon^*_i = \hat{\rho} \varepsilon^*_{i-1} + \varepsilon^*_i$ is iteratively applied to obtain the values $\varepsilon^*_i$ for $1 \leq i \leq n$.

Step 6: The bootstrap sample $(X, Y^*)$ is obtained by means of

$$Y^* = X \hat{\theta}_{\text{GMD}} + \varepsilon^*. \quad (7)$$

Step 7: With this bootstrap sample the GMD estimator $\hat{\theta}_{\text{GMD}}$ is computed. Thus,

$$\hat{\theta}_{\text{GMD}} = (X' \hat{A}^{-1} X)^{-1} (X' \hat{P} W_g \hat{P}^* Y^*), \quad (8)$$

where the matrices $\hat{P}^*$, $\hat{A}^{-1}$ and $W_g$ are analogous to $\hat{P}$, $\hat{A}$ and $W_g$. Nevertheless, they are obtained from bootstrap sample $(X, Y^*)$ by using the following estimator of $\rho^*$:

$$\rho^* = \frac{\sum_{i=1}^{n-1} \varepsilon^*_i \varepsilon^*_{i+1}}{\sum_{i=1}^{n} \varepsilon^*_i^2}. \quad (9)$$

Step 8: For a large value of $B$, Steps 5–7 are repeated $B$ times to obtain the bootstrap replications of the estimator: $\{\hat{\theta}_{\text{GMD},1}, \ldots, \hat{\theta}_{\text{GMD},B}\}$.

Comment 1. The computation of the $\hat{\theta}_{\text{GMD}}$ estimator and the described resampling mechanism can be generalized to linear models with random errors that follow an ARMA-dependence type. This is conceptually straightforward, but has the drawback of having to estimate a variance–covariance matrix that depends on more parameters.

Comment 2. In order to illustrate that for finite samples the bootstrap distribution is a better approximation to the distribution of $\hat{\theta}_{\text{GMD}}$ than the asymptotic distribution of the estimator, we have performed the following simulation study.

Let us consider the simple regression model

$$Y_i = \theta x_i + \varepsilon_i, \quad (10)$$

where $x_i = i/n$, $i = 1, \ldots, n$ and the observational errors, $\varepsilon_i$, follow an AR(1) model with marginal distribution $N(0, \sigma^2_e)$. 
Theorem 6 in Vilar-Fernández and González-Manteiga (2000) establishes the asymptotic normality of the \( \hat{\theta}_{\text{GMD}} \) estimator under suitable conditions. In particular, the asymptotic distribution of \( \hat{\theta}_{\text{GMD}} \) is shown to be \( N(\theta, 3\sigma^2_e / n(1-\rho)) \) when the data are generated from (10). This result has permitted us to compare the asymptotic density of \( \hat{\theta}_{\text{GMD}} \) with the nonparametric estimations of both the density function of \( \hat{\theta}_{\text{GMD}} \) and the density function \( \hat{\theta}^\star_{\text{GMD}} \). In order to do this, 1000 random samples of size \( n = 200 \) were drawn from model (10) with parameters \( \theta = 1, \sigma^2_e = 2 \) and the values \( \rho = 0.90 \) and \(-0.90\). The \( \hat{\theta}_{\text{GMD}} \) estimator was computed for each sample in such a way that the local linear regression was employed for the two smoothings, so that the Epanechnikov kernel was used and the bandwidths \( h \) and \( g \) were empirically chosen. From these 1000 values, the density function of \( \hat{\theta}_{\text{GMD}} - \theta \) was estimated by means of the Rosenblatt–Parzen kernel estimator. On the other hand, from each one of the simulated samples, a resample and the corresponding estimator \( \hat{\theta}^\star_{\text{GMD}} \) were obtained following the algorithm presented at the beginning of this section. Finally, with this set of 1000 values of \( \hat{\theta}^\star_{\text{GMD}} - \hat{\theta}_{\text{GMD}} \), the associated density function was estimated by using the Rosenblatt–Parzen kernel estimator again.

For the value \( \rho = 0.9 \), the graphs of the asymptotic density of \( \hat{\theta}_{\text{GMD}} - \theta \), that is \( N(0, 1.732^2) \), and of the estimated densities of \( \hat{\theta}_{\text{GMD}} - \theta \) (with standard deviation \( \sigma_{\theta} = 0.715 \)) and \( \hat{\theta}^\star_{\text{GMD}} - \hat{\theta}_{\text{GMD}} \) (with standard deviation \( \sigma_{\theta^\star} = 0.613 \)) have been plotted in Fig. 1. The analogous graphs when \( \rho = -0.9 \) are shown in Fig. 2 (in this case...
the asymptotic density is $N(0; 0.0912^2)$ and $\sigma_{\hat{\theta}}$ and $\sigma_{\hat{\theta}^*}$ take the values 0.0425 and 0.0428, respectively).

The combined analysis of both figures allows to notice that the bootstrap density is clearly closer to the estimated density than the normal approximation for large values of $\rho$ (either positive or negative). Further simulations have also confirmed the better behavior of the bootstrap density with respect to the asymptotic one for small values of $\rho$, although the improvement is less marked in this case. Finally, both the bootstrap and the asymptotic densities are very similar independently of the value considered for $\rho$ when very large sample sizes are taken.

**Comment 3.** One of the main applications of the bootstrap methodology consist of obtaining accurate confidence intervals for the parameter vector $\theta$, which can be done by using the percentile method or the percentile-t method (see González-Manteiga et al. (1994), for a description of these methods), among others. To illustrate this issue, our simulation study was extended as follows.

Random samples of size $n = 100$ were drawn from model (10) by considering two possible distributions for $e_i$: the normal and uniform distributions. The parameter $\theta = 1$ and two different values $\rho = 0.6$ and 0.9 were considered. For the variance of the errors we have taken $\sigma^2 = 2$ and $\sigma^2 = 4$. In every case, the number of simulated samples was
Table 1
Coverage percentages (CP), mean lengths and standard deviations (s.d. (l)) for the confidence intervals computed from the GMD estimator, $\hat{\theta}_{GMD}$, and from the GLS estimator, $\hat{\theta}_{GLS}$, when the error has a normal distribution.

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma^2 = 2$, $\rho = 0.6$</th>
<th></th>
<th></th>
<th>$\sigma^2 = 2$, $\rho = 0.9$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>Mean (l)</td>
<td>s.d. (l)</td>
<td>CP</td>
<td>Mean (l)</td>
<td>s.d. (l)</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, as.dist</td>
<td>94.60</td>
<td>1.918</td>
<td>0.557</td>
<td>95.60</td>
<td>5.945</td>
<td>3.672</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Perc.</td>
<td>88.70</td>
<td>1.468</td>
<td>0.326</td>
<td>78.60</td>
<td>2.287</td>
<td>0.820</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Per-t</td>
<td>89.10</td>
<td>1.499</td>
<td>0.320</td>
<td>84.10</td>
<td>2.759</td>
<td>1.229</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, SP-t</td>
<td>89.30</td>
<td>1.498</td>
<td>0.320</td>
<td>83.60</td>
<td>2.756</td>
<td>1.228</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, as.dist</td>
<td>91.80</td>
<td>1.913</td>
<td>0.555</td>
<td>93.70</td>
<td>5.869</td>
<td>3.604</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, SP-t</td>
<td>87.90</td>
<td>1.484</td>
<td>0.322</td>
<td>78.90</td>
<td>2.730</td>
<td>1.209</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma^2 = 4$, $\rho = 0.6$</th>
<th></th>
<th></th>
<th>$\sigma^2 = 4$, $\rho = 0.9$</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>Mean (l)</td>
<td>s.d. (l)</td>
<td>CP</td>
<td>Mean (l)</td>
<td>s.d. (l)</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, as.dist</td>
<td>96.80</td>
<td>2.683</td>
<td>0.764</td>
<td>98.00</td>
<td>8.776</td>
<td>5.741</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Perc.</td>
<td>91.80</td>
<td>2.060</td>
<td>0.440</td>
<td>81.90</td>
<td>3.317</td>
<td>1.175</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Per-t</td>
<td>92.40</td>
<td>2.104</td>
<td>0.434</td>
<td>85.90</td>
<td>4.015</td>
<td>1.857</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, SP-t</td>
<td>92.40</td>
<td>2.102</td>
<td>0.434</td>
<td>85.70</td>
<td>4.009</td>
<td>1.853</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, as.dist</td>
<td>94.30</td>
<td>2.676</td>
<td>0.761</td>
<td>95.50</td>
<td>8.670</td>
<td>5.661</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, SP-t</td>
<td>87.90</td>
<td>2.084</td>
<td>0.436</td>
<td>81.20</td>
<td>3.976</td>
<td>1.834</td>
</tr>
</tbody>
</table>

$N = 1000$ and, for each one of these samples, the bootstrap distribution of $\hat{\theta}_{GMD}^{*}$ was obtained from $B = 1000$ bootstrap replicates.

By using several methods, two-sided confidence intervals with coverage level $1 - \alpha = 0.90$ were derived. First, the asymptotic distribution of the $\hat{\theta}_{GMD}$ estimator was employed after replacing the parameters $\sigma_\varepsilon$ and $\rho$ by their respective estimations $\hat{\sigma}_\varepsilon$ and $\hat{\rho}$, which were computed from the residuals $\hat{\varepsilon}_i$.

Secondly, three bootstrap methods were used: the percentile method and two different versions of the percentile-$t$ method, as stated in Hall (1986). The percentile method considers the ordered values $\hat{\theta}_{GMD,(1)}^{*} \leq \hat{\theta}_{GMD,(2)}^{*} \leq \cdots \leq \hat{\theta}_{GMD,(B)}^{*}$ and then approximates the confidence interval by $(\hat{\theta}_{GMD,(r)}^{*}, \hat{\theta}_{GMD,(s)}^{*})$, with $r = [B\alpha/2]$ and $s = [B(1-\alpha)/2]$. The two versions of the percentile-$t$ method are: the equal tails percentile-$t$ (Per-$t$ in tables), based on the percentile method applied to the statistic $\sqrt{n}(\hat{\theta}_{GMD}^{*} - \hat{\theta}_{GMD})/\sigma_{\theta}^{*}$, and the symmetrized percentile-$t$ (SP-$t$ in tables), where one has to proceed in the same way but considering the statistic $|\sqrt{n}(\hat{\theta}_{GMD}^{*} - \hat{\theta}_{GMD})/\sigma_{\theta}^{*}|$. Thus, confidence intervals from $\hat{\theta}_{GMD}$ estimator were obtained by means of the three described bootstrap methods.

Finally, by employing the asymptotic distribution and the symmetrized percentile-$t$ method the confidence intervals from $\hat{\theta}_{GLS}$ estimator were also obtained. This was done to compare both estimators.

Table 1 shows the coverage percentages, mean lengths and standard deviations of the confidence intervals when the error has a normal distribution, while Table 2 shows the results when the error has a uniform distribution.
Table 2

Coverage percentages (CP), mean lengths and standard deviations (s.d. (l)) for the confidence intervals computed from the GMD estimator, $\hat{\theta}_{GMD}$, and from the GLS estimator, $\hat{\theta}_{GLS}$, when the error has a uniform distribution

<table>
<thead>
<tr>
<th>Method</th>
<th>$\sigma^2 = 2$, $\rho = 0.6$</th>
<th>$\sigma^2 = 2$, $\rho = 0.9$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CP</td>
<td>Mean (l)</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, as.dist</td>
<td>95.30</td>
<td>1.876</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Perc.</td>
<td>91.70</td>
<td>1.452</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, Per-t</td>
<td>91.90</td>
<td>1.475</td>
</tr>
<tr>
<td>$\hat{\theta}_{GMD}$, SP-t</td>
<td>91.40</td>
<td>1.474</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, as.dist.</td>
<td>93.40</td>
<td>1.871</td>
</tr>
<tr>
<td>$\hat{\theta}_{GLS}$, SP-t</td>
<td>87.30</td>
<td>1.463</td>
</tr>
<tr>
<td>$\sigma^2 = 4$, $\rho = 0.6$</td>
<td>94.50</td>
<td>2.716</td>
</tr>
<tr>
<td>$\sigma^2 = 4$, $\rho = 0.9$</td>
<td>92.20</td>
<td>2.708</td>
</tr>
</tbody>
</table>

From the numerical outputs presented in both tables, we conclude that the bootstrap methods lead to better results than the normal approximation. In addition to a level of coverage nearest to the theoretical covering, the bootstrap intervals present much smaller mean lengths and bigger standard deviations than those corresponding to the asymptotic intervals. With regard to the three described bootstrap methods, the percentile method provides worse results than the percentile-t methods, while significant differences between the equal tails percentile-t and the symmetrized percentile-t method are not observed. Finally, the bootstrap intervals obtained from the $\hat{\theta}_{GMD}$ estimator are better than those obtained from the $\hat{\theta}_{GLS}$ estimator because they have bigger coverage level with the same mean length.

Comment 4. Another additional motivation to study the proposed bootstrap algorithm comes from its use to test the initial hypothesis that the regression function follows a linear model. To make clear this interesting application, let us assume that we are interested in testing the linearity hypothesis

$$H_0: m \in \{m_0(x) = x'\theta, \ x \in C\},$$

against the alternative hypothesis $H_1$ : “$m$ is a smooth function”. As the functional $\Psi(\theta)$ defined in (2) measures a distance between a nonparametric estimation of $m(x)$ and the parametric estimation assumed under $H_0$, this value may be used as a statistic for the proposed test. On the other hand, the asymptotic normality of Crâmer-von-Mises-type functional distance $\Psi(\theta) = d(\hat{m}, m_0)$ was obtained in González-Manteiga and Vilar-Fernández (1995) and the consistency of the bootstrap distribution of $\Psi(\theta)$
was proved in Vilar-Fernández and González-Manteiga (1996) under general conditions.

An alternative approach would be to transform the initial data according to the model given in (5), obtaining the uncorrelated data \((\hat{X}, \hat{Y}) = (\hat{P}X, \hat{P}Y)\) and then computing the distance \(\Psi(\theta)\) in terms of the transformed data and the \(\hat{\theta}_{\text{GMD}}\) estimator. That is, a new test can be constructed by considering the functional distance

\[
\hat{\Psi}(\hat{\theta}_{\text{GMD}}) = \frac{1}{n} ((W_g \hat{P}Y - \hat{P}X \hat{\theta}_{\text{GMD}})'(W_g \hat{P}Y - \hat{P}X \hat{\theta}_{\text{GMD}}))
\]

\[
= \frac{1}{n} (M_g'(I - \hat{V}) M_g),
\]

where \(M_g = W_g \hat{P}Y\) and \(\hat{V} = \hat{X}(\hat{X}'\hat{X})^{-1}\hat{X}'\) is the orthogonal projection matrix over the subspace of \(\mathbb{R}^n\) generated by the columns of \(\hat{X} = \hat{P}X\).

A bootstrap procedure can also be used to approximate the distribution of \(\hat{\Psi}(\hat{\theta}_{\text{GMD}})\) and to compute the critical test points. More precisely, the bootstrap replications of both the estimator and the discrepancy measure \(\hat{\Psi}(\hat{\theta}_{\text{GMD}}, i) = D_i^*, i = 1, \ldots, B,\) are first obtained in Step 8 of the proposed resampling mechanism. Then, with a significance level \(\alpha\) the critical region of the test is given by

\[
\hat{\Psi}(\hat{\theta}_{\text{GMD}}) > D_{\lfloor(1-\alpha)B\rfloor}^*,
\]

where \([\cdot]\) represents the integer part and \(\{D_1^*, D_2^*, \ldots, D_B^*\}\) is the ordered sample of \(\{D_i^*\}, i = 1, \ldots, B.\)

3. Theoretical results

This section is devoted to present two results (Theorem 3.1 and Corollary 3.1) which establish the asymptotic validity of the bootstrap procedure proposed in Section 2. In order to prove Theorem 3.1, some assumptions on the starting model (1) and on the two weights sequences employed to compute the \(\hat{\theta}_{\text{GMD}}\) estimator are listed below.

The following assumption is made for the design:

A.1. Two positive definite matrices \(V_0\) and \(V_1\) of orders \(p \times p\) exist such that

\[
\lim_{n \to \infty} \frac{1}{n} X'X = V_0 \quad \text{and} \quad \lim_{n \to \infty} \frac{1}{n} X'A^{-1}X = V_1.
\]

Concerning the weights \(\{\omega_{ni,h}\},\) it is assumed that:

A.2. (i) \(\sum_{i=1}^n |\omega_{ni,h}(x)| \leq L,\) for every \(n\) and for every \(x \in C,\) where \(L\) is a positive constant.

(ii) \(\sum_{i=1}^n \omega_{ni,h}(x) \to 1,\) as \(n \to \infty,\) for every \(x \in C.\)

(iii) \(\sum_{i=1}^n |\omega_{ni,h}(x)| 1_{\|x'-x\| > \varepsilon} \to 0,\) as \(n \to \infty,\) for every \(\varepsilon > 0\) and for every \(x \in C,\) where \(\|\|\|\) is any norm of \(\mathbb{R}^p\) and \(1_{\{\}}\) represents the indicator function.

(iv) \(\sup_{\varepsilon \leq i \leq n} \|\omega_{ni,h}(x)\| \leq O(n^{-\nu}),\) for some \(\nu > 0\) and for every \(x \in C.\)

Finally, the following assumption is made for the weights \(\{\omega_{ni,g}\}:\)
A.3. Let \( \{\delta_n\} \) be a sequence of real numbers with \( \delta_n > 0 \) and \( n\delta_n^2 \downarrow 0 \) as \( n \uparrow \infty \). Let us consider the terms
\[
\zeta_n(x) = \sum_{i=1}^{n} \omega_{ni,q}(x) - 1 \quad \text{and} \quad \eta_n(x, \hat{\delta}_n) = \sum_{i=1}^{n} |\omega_{ni,q}(x)| 1\{||x_i - x|| > \delta_n\}.
\]
Then
\[
\sqrt{n}\zeta_n(x) \rightarrow 0 \quad \text{and} \quad \sqrt{n}\eta_n(x, \delta_n) \rightarrow 0
\]
holds as \( n \rightarrow \infty \), for every \( x \in C \).

The previous conditions are not very restrictive and have been employed in other papers. In particular, Assumption A.1 was employed by Stute (1995) to prove that the distribution of the two-stage GLS estimator admits a bootstrap approximation. Assumption A.2 ensures the consistency of the \( \hat{\theta}_{\text{MD}} \) estimator and, therefore, the consistency of the estimator for \( \rho \) given in (4). Assumption A.2 is less restrictive than Assumption A.3, but both of them are fulfilled by most of the classical nonparametric regression estimators: Nadaraya-Watson, Gasser-Müller and local polynomial fitting. See, for instance, the monographs of Härdle (1990) and Fan and Gijbels (1996) to study the properties of these estimators.

Theorem 3.1. Under assumptions A.1–A.3, we have, with probability one, that:
\[
\sqrt{n}(\hat{\theta}_{\text{MDG}} - \hat{\theta}_{\text{MDG}}) \rightarrow^d N(0, \sigma^2 V_1^{-1}) \quad \text{as} \quad n \rightarrow \infty,
\]
where \( \rightarrow^d \) denotes the convergence in distribution underlying the bootstrap samples.

Under the same set of assumptions, Vilar-Fernández and González-Manteiga (2000) have established the asymptotic normality of the \( \hat{\theta}_{\text{GMD}} \) estimator. In their Theorem 6, it is stated that
\[
\sqrt{n}(\hat{\theta}_{\text{GMD}} - \theta) \rightarrow^d N(0, \sigma^2 V_1^{-1}) \quad \text{as} \quad n \rightarrow \infty.
\]
Convergences (13) and (14) lead to the following result, that is analogous to Theorem 1.1 in Stute (1995) concerning a bootstrap approximation of the GLS estimator.

Corollary 3.1. Under the assumptions of Theorem 3.1, we have, with probability one, that
\[
\sup_{x \in \mathbb{R}^p} |P(n^{1/2}(\hat{\theta}_{\text{GMD}} - \theta) \leq x) - P^*(n^{1/2}(\hat{\theta}_{\text{GMD}}^* - \hat{\theta}_{\text{GMD}}) \leq x)| \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty,
\]
where \( P^* \) denotes the probability distribution under the resampling.

4. Proofs

In order to prove Theorem 3.1, four previous lemmas must first be obtained. From now on the \( d_2 \) metric (Mallow’s metric, also called Wasserstein distance) will be used.
Such a metric is defined for probability measures $P$ and $Q$, with $\int |x|^2 \, dP < \infty$ and $\int |x|^2 \, dQ < \infty$, in the form

$$d_2(P, Q) = \inf \left( E|X - Y|^2 \right)^{1/2},$$

where the infimum is taken over pairs $(X, Y)$ of random variables, with $X$ and $Y$ distributed according to $P$ and $Q$, respectively. By means of Mallow’s metric it can be shown that $\hat{F}_n$ approximates to $F$ asymptotically, where $F$ is the distribution function of the noise process \( \{e_i\} \) and $\hat{F}_n$ is the empirical distribution obtained in Step 4 in Section 2.

**Lemma 4.1.** Under the assumptions of Theorem 3.1 we have, with probability one, that

$$d_2(F, \hat{F}_n) \to 0 \quad \text{as } n \to \infty. \quad (15)$$

**Proof.** Let us denote by $F'_n$ the empirical distribution function based on the error series $\{e'_i\}_{i=2}^n$, which is obtained as the $\hat{e}_i$ (Steps 2 and 3 in Section 2), but from the unobservable residuals $\{L_{SI_i}\}_{i=1}^n$. So, we have

$$d_2(\hat{F}_n, F) \leq d_2(\hat{F}_n, F'_n) + d_2(F'_n, F). \quad (16)$$

Moreover, from Theorem 3.1 in Kreiss and Franke (1992), we conclude that $d_2(F'_n, F) \to 0$ as $n \to \infty$ almost everywhere.

With regard to $d_2(\hat{F}_n, F'_n)$, let us denote by $P'$ the matrix similar to $\hat{P}$, obtained from vector $\varepsilon$. Then

$$d_2^2(\hat{F}_n, F'_n) \leq \frac{1}{n - 1} \sum_{i=2}^n (\hat{e}_i - e'_i)^2$$

$$= \frac{1}{n - 1} (\hat{P}_\varepsilon - P'_\varepsilon) \hat{Y} (\hat{P}_\varepsilon - P'_\varepsilon)$$

$$= \frac{1}{n - 1} (\hat{P}_\varepsilon - \varepsilon) \hat{Y} \hat{P} (\hat{P}_\varepsilon - \varepsilon) + \frac{2}{n - 1} (\hat{P}_\varepsilon - \varepsilon) \hat{Y} \hat{P}' (\hat{P} - P') \varepsilon$$

$$+ \frac{2}{n - 1} (\hat{P} - P') \hat{Y} (\hat{P} - P') \varepsilon$$

$$= \frac{1}{n - 1} (\theta - \hat{\theta}_{GMD})' X' \hat{A}^{-1} X (\theta - \hat{\theta}_{GMD}) + \frac{2}{n - 1} (\theta - \hat{\theta}_{GMD})' X' \hat{P}' \hat{P}_\varepsilon$$

$$- \frac{2}{n - 1} (\theta - \hat{\theta}_{GMD})' X' \hat{P}' P'_\varepsilon + \frac{2}{n - 1} (\theta - \hat{\theta}_G) \hat{P} - (\theta - \hat{\theta}_G) \hat{P} - (\theta - \hat{\theta}_G) \hat{P}.'

Now, Assumption A.1 together with the strong consistency of estimators $\hat{\rho}$ and $\hat{\theta}_{GMD}$ (Theorem 3 in Vilar-Fernández and González-Manteiga, 2000) lead to the convergence to zero of $d_2(\hat{F}_n, F'_n)$ and thus (15) is stated.

With similar arguments to those employed by Stute (1995), the following result is obtained.
Lemma 4.2. Under the assumptions of Theorem 3.1 we have, with probability one, that
\[ \rho^* - \hat{\rho} \to 0 \quad \text{as } n \to \infty, \tag{17} \]
where \( \rho^* = \sum_{i=1}^{n} \bar{\varepsilon}_i^* \bar{\varepsilon}_{i+1}^*/\sum_{i=1}^{n} \bar{\varepsilon}_i^2 \) being \( \bar{\varepsilon}_i^* \) those values computed in Step 5 in Section 2.

Proof. It is sufficient to establish as \( n \to \infty \) the following two convergences:
\[ \left( \frac{1}{n} \sum_{i=1}^{n-1} \bar{\varepsilon}_{i+1}^* \right) \left( \frac{1-\hat{\rho}^2}{\hat{\rho}} \right) \to \sigma_e^2, \tag{18} \]
\[ \left( \frac{1}{n} \sum_{i=1}^{n} \bar{\varepsilon}_i^2 \right) (1-\hat{\rho}^2) \to \sigma_e^2. \tag{19} \]

The proofs of (18) and (19) are obtained in a similar way, so that only the arguments for the proof of the convergence (18) are included below.

As the sequence \( \{ \bar{\varepsilon}_i^* \}_{i=1}^{n} \) satisfies the recursive expression \( \bar{\varepsilon}_i^* = \hat{\rho} \bar{\varepsilon}_{i-1}^* + \varepsilon_i^* \), we can write
\[ \left( \frac{1-\hat{\rho}^2}{\hat{\rho}} \right) \frac{1}{n} \sum_{i=1}^{n-1} \bar{\varepsilon}_{i+1}^* \bar{\varepsilon}_i^* = \frac{\hat{\rho}}{n} (\bar{\varepsilon}_1^* \bar{\varepsilon}_n^* - \bar{\varepsilon}_0^* \bar{\varepsilon}_{n-1}^*) + \frac{1}{n} \sum_{i=1}^{n-1} \bar{\varepsilon}_{i-1}^* \bar{\varepsilon}_{i+1}^* \]
\[ + \frac{\hat{\rho}}{n} \sum_{i=1}^{n-1} \bar{\varepsilon}_{i-1}^* e_i^* + \frac{1}{n} \sum_{i=1}^{n} \bar{\varepsilon}_i^2 + \frac{1}{n \hat{\rho}} \sum_{i=1}^{n-1} e_i^* \sum_{i=1}^{n} e_{i+1}^* \tag{20} \]

From the analysis of the right-hand side in expression (20), we observe that the first term converges to zero in probability and the fourth one converges in probability to the variance of \( \varepsilon_i^* \) by the weak law of large numbers and, in consequence, this term tends to \( \sigma_e^2 \) by Lemma 4.1. As the remaining sums consist of uncorrelated terms, we obtain their convergence to zero by using Lemma 4.1 again.

Lemma 4.3. Under the assumptions of Theorem 3.1, the following convergences are verified:
\[ \frac{1}{\sqrt{n}} X^t \varepsilon^* = O_p(1), \tag{21} \]
\[ \frac{1}{n} X^t (\hat{A}^{*-1} - \hat{A}^{-1}) X \to 0 \quad \text{almost sure}, \tag{22} \]
\[ \frac{1}{\sqrt{n}} X^t (\hat{A}^{*-1} - \hat{A}^{-1}) \varepsilon^* \to 0 \quad \text{almost sure}. \tag{23} \]

Lemma 4.3 is analogous to Lemma 7 in Vilar-Fernández and González-Manteiga (2000), but this time for the bootstrap variables. For this reason, its proof is omitted here.
Lemma 4.4. Under Assumptions A.1 and A.3, we have
\[
\sqrt{n}X'W_gX = \sqrt{n}X'X + o_p(1). \tag{24}
\]

Proof. For any \(\delta_n > 0\) and for \(i = 1, \ldots, n\), we can write
\[
\sum_{j=1}^{n} x_i x_j \omega_{nj,g}(x_i) = \sum_{j=1}^{n} x_i (x_j' - x_i') I\{\|x_j - x_i\| \leq \delta_n\} \omega_{nj,g}(x_i) \\
+ \sum_{j=1}^{n} x_i (x_j' - x_i') I\{\|x_j - x_i\| > \delta_n\} \omega_{nj,g}(x_i) \\
+ \left( \sum_{j=1}^{n} \omega_{nj,g}(x_i) - 1 \right) x_i x_i' \\
\leq C_1 \delta_n (\xi_n(x_i) + 1) + C_2 \eta_n(x_i, \delta_n) + C_3 \xi_n(x_i) + x_i x_i',
\]
where \(C_k\) \((k = 1, 2, 3)\) are positive constants. Now, by choosing \(\delta_n = o(n^{-1/2})\) and using Assumptions A.1 and A.3, the identity (24) is concluded and the lemma is proved.

Proof of Theorem 3.1. Let \(\hat{\theta}^*\) denote the estimator
\[
\hat{\theta}^* = (X'\hat{A}^{-1}X)^{-1}X'\hat{P}^t W_g \hat{P} Y^*.
\]

Let
\[
n^{1/2}(\hat{\theta}_{GMD}^* - \hat{\theta}_{GMD}) = n^{1/2}(\hat{\theta}_{GMD}^* - \hat{\theta}^*) + n^{1/2}(\hat{\theta}^* - \hat{\theta}_{GMD}). \tag{25}
\]

By considering (7) and (8), the first term in the right-hand side of (25) can be split into two parts
\[
n^{1/2}(\hat{\theta}_{GMD}^* - \hat{\theta}^*) = A_D + A_R, \tag{26}
\]
where
\[
A_D = n^{1/2}(X'\hat{A}^* X)^{-1}X'\hat{P}^t W_g \hat{P} X \hat{\theta}_{GMD} - (X'\hat{A}^{-1}X)^{-1}X'\hat{P}^t W_g \hat{P} X \hat{\theta}_{GMD})
\]
and
\[
A_R = n^{1/2}(X'\hat{A}^* X)^{-1}X'\hat{P}^t W_g \hat{P} \varepsilon^* - (X'\hat{A}^{-1}X)^{-1}X'\hat{P}^t W_g \hat{P} \varepsilon^*).
\]

First, we focus on the random part \(A_R\) in (26). We have that
\[
A_R = n^{1/2}(X'\hat{A}^* X)^{-1} - (X'\hat{A}^{-1}X)^{-1})X'\hat{P}^t W_g \hat{P} \varepsilon^* \\
+ n^{1/2}(X'\hat{A}^* X)^{-1}(X'\hat{P}^t W_g \hat{P} \varepsilon^* - X'\hat{P}^t W_g \hat{P} \varepsilon^*) \\
= [(n^{-1}X'\hat{A}^* X) - (n^{-1}X'\hat{A}^* X)]X + n^{-1}X'\hat{A}^{-1}X)^{-1} \times n^{-1/2}X'\hat{P}^t W_g \hat{P} \varepsilon^* \\
+ n^{-1}X'\hat{A}^{-1}X)^{-1}n^{-1/2}(X'\hat{P}^t W_g \hat{P} \varepsilon^* - X'\hat{P}^t W_g \hat{P} \varepsilon^*). \tag{27}
\]
By using Assumption A.1 and convergence (22) in Lemma 4.3, we obtain that
\[ [(n^{-1}X'(A^{-1} - \hat{A}^{-1})X + n^{-1}X'\hat{A}^{-1}X)^{-1} - (n^{-1}X'\hat{A}^{-1}X)^{-1}] = o_P(1). \] (28)

In addition, by Lemma 4.4 we have
\[ n^{1/2}X'\hat{P}^*W'g\hat{P}^*\varepsilon = n^{1/2}X'\hat{A}^{-1}X + o_P(1). \] (29)

From (29) and (21) in Lemma 4.3, it follows that
\[ n^{-1/2}X'\hat{P}^*W'g\hat{P}^*\varepsilon = o_P(1). \] (30)

The convergence to zero of \( n^{-1/2}(X'\hat{P}^*W'g\hat{P}^*\varepsilon - X'\hat{P}^*W_g\hat{P}^*\varepsilon) \), derived in a similar way by using expression (23) in Lemma 4.3. This fact together with (28) and (30) lead to establish, with probability one, the convergence to zero of the random part, \( A_R \), in (27).

Analogous arguments (using Assumption A.1, Lemmas 4.3 and 4.4 again) are valid to show that \( A_D = o_P(1) \), so that we can conclude that the term \( n^{1/2}(\hat{\theta}^* - \hat{\theta}_{\text{GMD}}) \) in (25) converges to zero with probability one.

Now, we study the asymptotic distribution of the second term in (25): \( n^{1/2}(|\hat{\theta}^* - \hat{\theta}_{\text{GMD}}|) \). Lemma 4.4 leads to
\[ n^{1/2}(\hat{\theta}_{\text{GMD}} - \theta) \overset{d}{\rightarrow} N(0, \sigma^2_\varepsilon \varepsilon^{-1}) \quad \text{as} \quad n \to \infty, \] where
\[ n^{1/2}(\hat{\theta}_{\text{GMD}} - \theta) = n^{1/2}((X'\hat{A}^{-1}X)^{-1}X'\hat{P}^*W_g\hat{P}^*\varepsilon + o_P(1)). \] (32)

So, taking into account Assumption A.1, to prove Theorem 3.1, it would suffice to show that the random vectors \( \tilde{\varepsilon}^* = n^{-1/2}X'\hat{P}^*W_g\hat{P}^*\varepsilon \) and \( \tilde{\varepsilon} = n^{-1/2}X'\hat{P}^*W_g\hat{P}^*\varepsilon \), with distribution functions \( \Phi^* \) and \( \Phi \), respectively, verify that \( d_2(\Phi^*, \Phi) \) converges to zero with probability one.

As in the proof of Lemma 4.1, we assume that the unobserved residuals \( \{e_i\}_{i=1}^n \) are known, so that Steps 2–5 of the bootstrap procedure in Section 2 can be performed starting from the values of \( \epsilon_i \). In such a case, the corresponding outputs will be denoted by \( \rho', P', \Lambda'^{-1}, \{\epsilon'_i\}_{i=1}^n, F'_n, \{(\epsilon'_i^*)^n_{i=1}, \{(\epsilon'_i^*)^n_{i=1}, Y'^* \) and \( W'_g \).

Let \( \Phi' \) denote the distribution of the random vector \( \tilde{\varepsilon}' = n^{-1/2}X'P'^*W_gP'\varepsilon^* \). Then
\[ d_2(\Phi^*, \Phi') \leq d_2(\Phi^*, \Phi') + d_2(\Phi', \Phi). \] (33)

A similar reasoning scheme to the one carried out in the proof of Theorem 4.1 in Kreiss and Franke (1992) leads to the convergence to zero of the Mallows distance \( d_2(\Phi', \Phi) \). Concerning the last term in (33), we have
\[ d_2(\Phi^*, \Phi') = \inf \|n^{-1/2}X'\hat{P}^*W_g\hat{P}^*\varepsilon - n^{-1/2}X'P'^*W_gP'\varepsilon^* \|, \] (34)
where the \( i \)th components of \( \varepsilon^* \) and \( \varepsilon'^* \) are given by \( \sum_{j=0}^{\infty} \hat{\rho}^j e^*_{i-j} \) and \( \sum_{j=0}^{\infty} \rho'^j e'^*_{i-j} \), respectively, the infimum is taken over independent and identically distributed pairs.
\((e^*_i, e'_i)\) with marginal distributions \(\hat{F}_n\) and \(F'_n\), and \(\| \cdot \|\) is the \(L^2\) norm conditioned to the values \(e_1, e_2, \ldots, e_n\).

Note that the infimum in (34) can be bounded by
\[
\inf \{ \| n^{-1/2} X^t \hat{P}^t W_g \hat{P}(\zeta^* - \zeta'^*) \| + \| n^{-1/2} (X^t \hat{P}^t W_g \hat{P} - X^t P'' W_g P') \zeta'^* \| \}. \tag{35}
\]

Once again, by using Assumptions A.1, A.3 and Lemmas 4.3 and 4.4 as they were used to establish the convergence of \(\Delta_R\) in (26), we obtain that
\[
\| n^{-1/2} (X^t \hat{P}^t W_g \hat{P} - X^t P'' W_g P') \zeta'^* \| \simeq n^{-1/2} (X' (\hat{\Lambda}^{-1} - \Lambda'^{-1}) \zeta'^*) = o_p(1). \tag{36}
\]

On the other hand, if \(\{ \chi_{l,i} \} \) denote the \(p \times n\) matrix \(X^t \hat{P}^t W_g \hat{P}\), we find
\[
\| n^{-1/2} X^t \hat{P}^t W_g \hat{P}(\zeta^* - \zeta'^*) \|
= \| n^{-1/2} \begin{pmatrix}
\sum_{i=1}^n \chi_{1,i} \left( \sum_{j=0}^\infty (\hat{P}^t e^*_i - \rho^t e'_i) \right)
\vdots
\sum_{i=1}^n \chi_{p,i} \left( \sum_{j=0}^\infty (\hat{P}^t e^*_i - \rho^t e'_i) \right)
\end{pmatrix}
\| \\
= \| n^{-1/2} \begin{pmatrix}
\sum_{i=1}^n \chi_{1,i} \left( \sum_{j=0}^\infty (\hat{P}^t e^*_i - e'_i) \right)
\vdots
\sum_{i=1}^n \chi_{p,i} \left( \sum_{j=0}^\infty (\hat{P}^t e^*_i - e'_i) \right)
\end{pmatrix}
\| \\
+ \| n^{-1/2} \begin{pmatrix}
\sum_{i=1}^n \chi_{1,i} \left( \sum_{j=0}^\infty (\hat{P}^t - \rho^t) e'_i \right)
\vdots
\sum_{i=1}^n \chi_{p,i} \left( \sum_{j=0}^\infty (\hat{P}^t - \rho^t) e'_i \right)
\end{pmatrix}
\|
= S_1 + S_2.
\]

Since \(e^*_{i-j} - e'_i\) are independent and identically distributed, we have
\[
S^2_1 = \frac{1}{n} \sum_{i=1}^p E \left[ \left( \sum_{i=1}^n \chi_{1,i} \left( \sum_{j=0}^\infty (\hat{P}^t e^*_i - e'_i) \right) \right)^2 \right]
= \sum_{i=1}^p \left( \frac{1}{n} \sum_{i=1}^n \chi^2_{1,i} \left( \sum_{j=0}^\infty \rho^t \right) \right) E(e^*_{i-j} - e'_i)^2 \leq C d_2^2(\hat{F}_n, F'_n) = o_p(1),
\]
where we have employed Assumption A.1 and Lemma 4.1.
Analogously, we obtain

\[
S_2^2 = \sum_{l=1}^{p} \frac{1}{n} \left[ \sum_{i=1}^{n} \left( \sum_{j=0}^{\infty} (\hat{\rho}^j - \rho^j) e_{i-j}^* \right) \right]^2 \\
= \sum_{l=1}^{p} \left( \frac{1}{n} \sum_{i=1}^{n} \hat{\gamma}_l^2 \right) \left( \sum_{j=0}^{\infty} (\hat{\rho}^j - \rho^j)^2 \right) E \left( e_{i-j}^2 \right) \\
\leq C \sum_{j=0}^{\infty} |\hat{\rho}^j - \rho^j|^2 \leq C|\hat{\rho} - \rho|^2 \sum_{j=0}^{\infty} \nu^2j = o_P(1),
\]

where \( \nu \in (0,1) \) and we have used \( \hat{\rho} \) and \( \rho^j \) as consistent estimators of \( \rho \). Thus, the proof of Theorem 3.1 is completed.

References


