Supplementary Material to *Strongly-consistent autoregressive predictors in abstract Banach spaces*

M. D. Ruiz-Medina†and J. Álvarez-Liébana†

†*Department of Statistics and O. R. (mruiz@ugr.es, javialvaliebana@ugr.es)*
Faculty of Sciences, Univesity of Granada
Campus Fuente Nueva s/n
18071 Granada, Spain

Abstract

This document provides the Supplementary Material to the paper entitled *Strongly-consistent autoregressive predictors in abstract Banach spaces.* Specifically, a simulation study is undertaken to illustrate the results derived, on strong consistency of functional predictors, in abstract Banach spaces, from the ARB(1) framework. The results are also illustrated in the case of discretely observed functional data.

Keywords: ARB(1) processes, Banach spaces, continuous embeddings, functional plug-in predictors, strongly-consistent estimators

1. Simulation study

The sequences $\{F_{\mathbf{m}}\}$ and $\{t_{\mathbf{m}}\}$, appearing in the formulation of Lemma 1, are defined in equations (52)–(54) of Section 6 of the manuscript. They define an orthogonal wavelet basis in $L^2([0,1])$, the space H in Lemma 2. In the simulation study undertaken, we have selected the Daubechies wavelet basis of order $N = 10$, providing The norm $\|\cdot\|_B$ of our Banach space $B = B_{\infty,\infty}([0,1])$ is then computed from equation (55) in the manuscript as follows: For a coarser resolution level J up to a resolution given by the truncation parameter M: For every $f \in B$,

$$\|f\|_B \;=\; \sup\left\{\left|\alpha_{J,k}^f\right|,\ k=0,\ldots,2^{J-1}; \left|\beta_{j,k}^f\right|,\ k=0,\ldots,2^j-1;\ j=J,\ldots,M\right\} \tag{1}$$

where

$$\begin{aligned}\alpha_{J,k}^{f} &= \int_0^1 f(t)\overline{\varphi_{J,k}(t)}dt, \quad k=0,\ldots,2^J-1,\\ \beta_{j,k}^{f} &= \int_0^1 f(t)\overline{\psi_{j,k}(t)}dt \quad k=0,\ldots,2^j-1, \quad j\geq J.\end{aligned}$$

Thus, equation (1) corresponds to the choice $B = B^0_{\infty,\infty}([0,1])$, when resolution level M is fixed for truncation. Therefore, $B^\star = B^0_{1,1}([0,1])$ is considered with the truncated norm

$$\|g\|_{B^\star} = \sum_{k=0}^{2^J-1}\left|\alpha_{J,k}^{g}\right| + \sum_{j=J}^{M}\sum_{k=0}^{2^j-1}\left|\beta_{j,k}^{g}\right|, \quad \forall g\in B^\star, \tag{2}$$

where $\{\alpha_{J,k}^{g}\}$ and $\{\beta_{j,k}^{g}\}$ are the respective father and mother wavelet coefficients of function g (see also equation (56) of the manuscript). Furthermore, as given in Section 6 of the manuscript (see also equation (52)), $\widetilde{H}^\star = H_2^\beta([0,1]) = B_{2,2}^\beta([0,1])$, and $\widetilde{H} = H_2^{-\beta}([0,1]) = B_{2,2}^{-\beta}([0,1])$, for $\beta > 1/2$. Since Daubechies wavelets of order $N = 10$ are selected as orthogonal wavelet basis, with $N = 10$ vanishing moments, according to [2, p. 271 and Lemma 2.1], and [3, p. 153], we have considered $J = 2$, and $M = \lceil \log_2(L/2)\rceil = 10$, for $L = 2^{11}$ nodes, in the discrete wavelet transform applied. In addition, value $\beta = 6/10 > 1/2$ has been tested, with $\gamma = 2\beta + \epsilon$, $\epsilon = 0.01$ (see definition of the extended version of operator C on $\widetilde{H} = H^{-\beta}([0,1])$, in Section 6, in equation (58)). The covariance kernel is now displayed in Figure 1 (see [6, pp. 119–140] and [7, p. 6]).

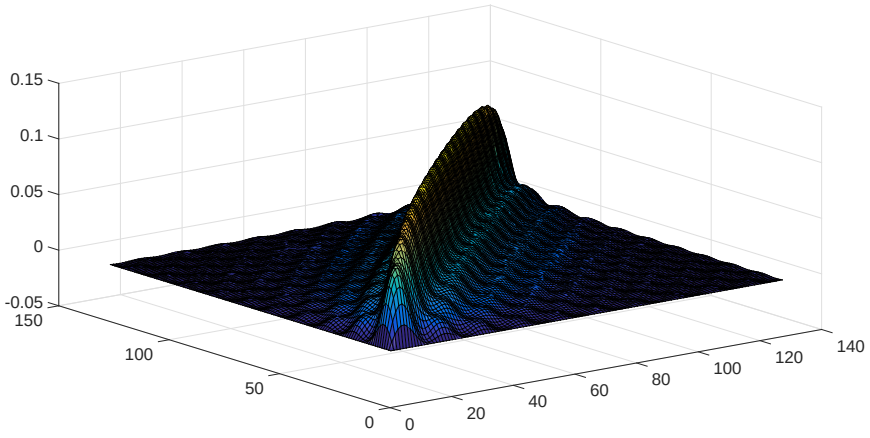

Figure 1: Covariance kernel defining C, generated with discretization step size $\Delta h = 0.0372$.

Under **Assumption A3**, operator ρ admits the following extended representation in $\widetilde{H} = H^{-\beta}([0,1])$, and in B :

$$\langle \rho(\phi_j), \phi_h \rangle_{H^{-\beta}([0,1])} = \begin{cases} (1+j)^{-1.5} & j = h \\ e^{-|j-h|/W} & j \neq h \end{cases},$$

Operator C_ε also admits, in this case, the following extended version in $\widetilde{H} = H^{-\beta}([0,1])$:

$$\langle C_\varepsilon(\phi_j), \phi_h \rangle_{H^{-\beta}([0,1])} = \begin{cases} C_j \left(1 - \rho_{j,j}^2\right) & j = h \\ e^{-|j-h|^2/W^2} & j \neq h \end{cases},$$

being $W = 0.4$.

1.1. Large-sample behaviour of the ARB(1) plug-in predictor

The ARB(1) process is generated with discretization step size $\Delta h = 0.0372$. The resulting functional values of ARB(1) process X are showed in Figure 2 at the times

$$t = [2500, 5000, 15000, 25000, 40000, 55000, 80000, 100000, 130000, 165000] .$$

In this section (but not in the next section), the generated discrete values are interpolated and smoothed, applying the *'cubicspline'* option in *'fit.m'*

MatLab function, with, as commented before, the number of nodes $L = 2^{11} = 2048$, then $M = 10$, and $\Delta\widetilde{h} = 0.0093$. In the following computations, $N = 250$ replications are generated for each functional sample size, and $k_n = \ln(n)$ has been tested.

The random initial condition X_0 has been generated from a truncated zero mean Gaussian distribution. Figure 3 illustrates the fact that **Assumption A1** holds, and Figure 4 is displayed to check **Assumption A2**.

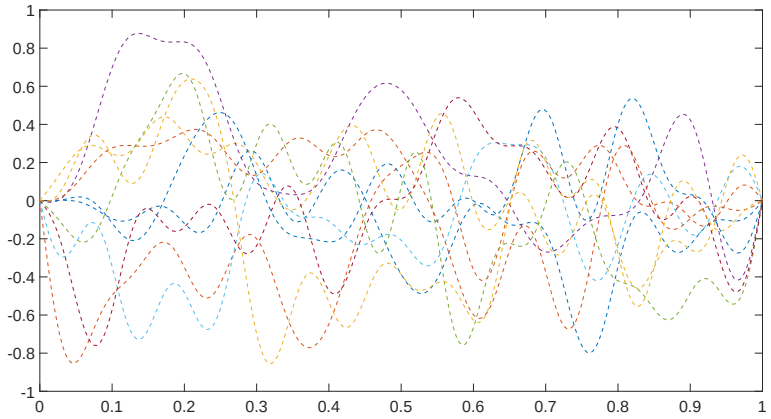

Figure 2: Functional values X_t, for $t(10^{-3}) = 2, 5, 5, 15, 25, 40, 55, 80, 100, 130, 165$ and discretization step size $\Delta h = 0.0372$.

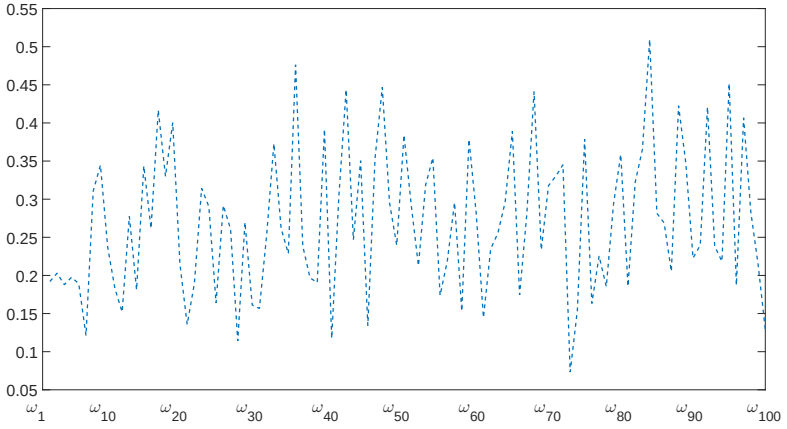

Figure 3: 100 values, $\|X_0(\omega_l)\|_B$, $l = 1, \ldots, 100$, (blue dotted line) are generated, for discretization step $\Delta h = 0.0372$.

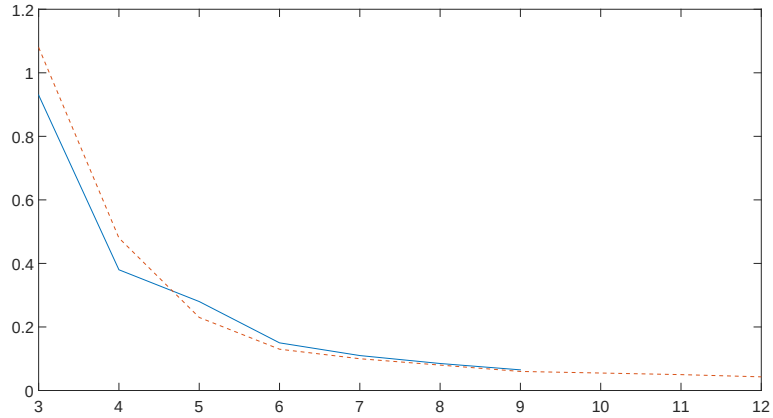

Figure 4: **Assumption A2** is checked for sample sizes $n_t = 35000$ (blue line) and $n_t = 395000$ (orange dotted line), displaying the decay rate of empirical eigenvalues $\{C_{n,j},\ j = 3, \ldots, k_n\}$, being $k_n = \lceil \ln(n) \rceil$.

Condition (39) in Theorem 1 has been checked as well (see Figure 5).

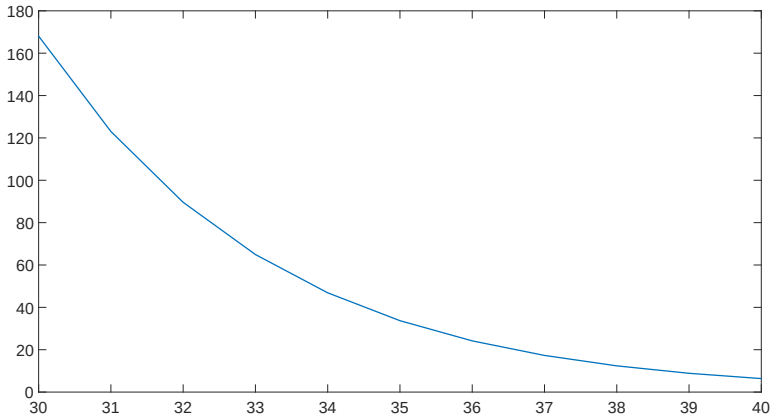

Figure 5: Values for $\left(k_n C_{k_n}^{-1} \sum_{j=1}^{k_n} a_j \right) \left(n^{1/2} \left(\ln(n) \right)^{-1/2} \right)^{-1}$, tested for truncation parameters $k_n = 30, \ldots, 40$, linked to sample sizes by the truncation rule $k_n = \ln(n)$.

To illustrate Theorem 1 and Corollary 1 in the paper, Table 1 displays the proportion of values of the random variable $\left\| \rho\left(X_{n_t}\right) - \widehat{X}_{n_t+1} \right\|_B$ that are

larger than the upper bound

$$\xi_{n_t} = \exp\left(\frac{-n_t}{C_{k_{n_t}}^{-2} k_{n_t}^2 \left(\sum_{j=1}^{k_{n_t}} a_j\right)^2}\right), \quad t = 1, \ldots, 10, \tag{3}$$

from the 250 values generated, for each functional sample size n_t, $t = 1, \ldots, 10$, reflected below.

Table 1: Proportion of simulations whose error B-norm is larger than the upper bound in equation (3). Truncation parameter $k_n = \ln(n)$, and $N = 250$ realizations have been considered, for each functional sample size.

n_t	
$n_1 = 2500$	$\frac{13}{250}$
$n_2 = 5000$	$\frac{11}{250}$
$n_3 = 15000$	$\frac{7}{250}$
$n_4 = 25000$	$\frac{4}{250}$
$n_5 = 40000$	$\frac{2}{250}$
$n_6 = 55000$	$\frac{1}{250}$
$n_7 = 80000$	0
$n_8 = 100000$	$\frac{1}{250}$
$n_9 = 130000$	0
$n_{10} = 165000$	0

Figure 6 below illustrates the asymptotic efficiency. The curve $n^{-1/4}$ is also displayed (red dotted line).

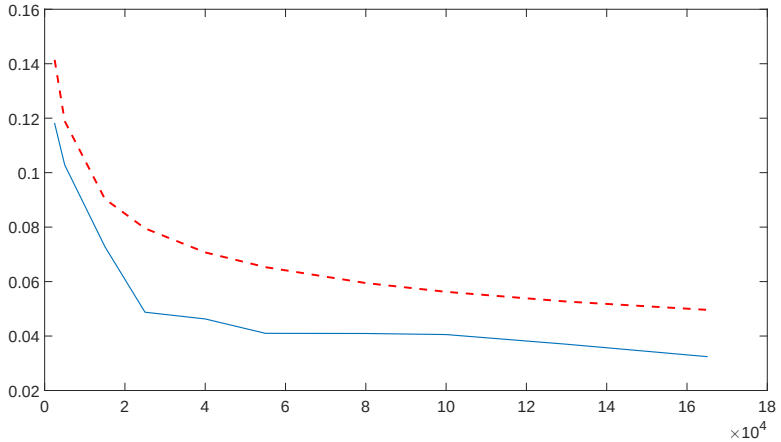

Figure 6: **Asymptotic efficiency**. Empirical mean-square error (blue solid line) $\mathrm{E}\left\{\left\|\rho\left(X_{n_t}\right)-\widehat{X}_{n_t+1}\right\|_B^2\right\}$, based on $N=250$ simulations. The curve $n^{-1/4}$ is also drawn (red dotted line).

1.2. Asymptotic behaviour of discretely observed ARB(1) processes

The results in Theorem 1 and Corollary 1 are now tested for different discretazitation step sizes:

$$\left\{\Delta h_r=\left(2^{8+r}-1\right)^{-1},\ r=1,\ldots,7\right\},\quad \Delta h_r\longrightarrow^{r\to\infty}0,$$

that is,

$$\begin{aligned}\Delta h_1 &= 1.96\left(10^{-3}\right), \quad \Delta h_2=9.78\left(10^{-4}\right),\\ \Delta h_3 &= 4.89\left(10^{-4}\right), \quad \Delta h_4=2.44\left(10^{-4}\right),\\ \Delta h_5 &= 1.22\left(10^{-4}\right), \quad \Delta h_6=6.10\left(10^{-5}\right),\\ \Delta h_7 &= 3.06\left(10^{-5}\right).\end{aligned}$$

Due to computational limitations involved in the smallest discretization step sizes, we restrict our attention here to the sample sizes

$$\left\{n_t=5000+10000\left(t-1\right),\ t=1,2,3\right\},$$

and $N=120$ realizations have been generated, for each functional sample size. The same nodes are considered as in the previous section, in the implementation of the discrete wavelet transform, without previous smoothing of the discretely generated data.

Tabla 2 displays the results obtained on the proportion of values, from the 120 generated values, $\left\|\rho\left(X_{n_t}^{h,r}\right)-\widehat{X}_{n_t+1}^{h,r}\right\|_B$, $h=1,\ldots,120$, that are larger than the upper bound (3), considering different discretization step sizes, for each sample size $\{n_t=5000+10000\,(t-1)\,,\ t=1,2,3\}$, and for the corresponding truncation orders $k_{n_t}=\ln(n_t)$, $t=1,2,3$.

Table 2: Proportions of simulations whose error B-norms are larger than the upper bound in (3), for sample sizes $n=[5000,\ 15000,\ 35000]$. Truncation parameter $k_n=\ln(n)$ has been considered. For each one of the functional sample sizes, the results displayed correspond to discretization step sizes $\left\{\Delta h_r=\left(2^{8+r}-1\right)^{-1},\ r=1,\ldots,7\right\}$. $N=120$ simulations are generated, for each sample and discretization step size.

	$n_1=5000$	$n_2=15000$	$n_3=35000$
$\Delta h_1=1.96\left(10^{-3}\right)$	$\frac{12}{120}$	$\frac{7}{120}$	$\frac{6}{120}$
$\Delta h_2=9.78\left(10^{-4}\right)$	$\frac{8}{120}$	$\frac{4}{120}$	$\frac{4}{120}$
$\Delta h_3=4.89\left(10^{-4}\right)$	$\frac{4}{120}$	$\frac{2}{120}$	$\frac{2}{120}$
$\Delta h_4=2.44\left(10^{-4}\right)$	$\frac{2}{120}$	$\frac{1}{120}$	$\frac{1}{120}$
$\Delta h_5=1.22\left(10^{-4}\right)$	$\frac{2}{120}$	$\frac{1}{120}$	0
$\Delta h_6=6.10\left(10^{-5}\right)$	$\frac{1}{120}$	0	0
$\Delta h_7=3.06\left(10^{-5}\right)$	$\frac{1}{120}$	0	0

Acknowledgments

This work was supported in part by project MTM2015–71839–P (co-funded by Feder funds), of the DGI, MINECO, Spain.

References

[1] R. Adams and J. Fournier (2003). *Sobolev spaces (2ed)*. Academic Press, Massachusetts.

[2] C. Angelini, D. De Canditiis and F. Leblanc (2003). Wavelet regression estimation in nonparametric mixed effect models. *J. Multivariate Anal.* **85**, 267–291.

[3] A. Antoniadis and T. Sapatinas (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.* **87**, 133–158.

[4] A. Cohen, I. Daubechies and P. Vial (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harm. Anal.* **1**, 54–81.

[5] I. Daubechies (1988). Orthonormal bases of compactly supported wavelets. *Comm. Pure and Appl. Math.* **41**, 909–996.

[6] R. Dautray and J. L. Lions (1990). *Mathematical Analysis and Numerical Methods for Science and Technology Volume 3: Spectral Theory and Applications*. Springer, New York.

[7] D. S. Grebenkov and B. T. Nguyen (2013). Geometrical structure of Laplacian eigenfunctions. *SIAM Review* **55**, 601–667.

[8] W. Hardle, G. Kerkyacharian, D. Picard and A. Tsybakov (1998). *Wavelets, approximation and statistical applications*. Springer-Verlag, New York.

[9] S. G. Mallat (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Inteligence* **11**, 674–693.

[10] Y. Meyer (1991). Ondelettes sur l'intervalle. *Rev. Mat. Iberoamericana* **7**, 115–133.

[11] H. Triebel (1983). *Theory of function spaces II*. Birkhauser, Basel.

[12] B. Vidakovic and P. Muller (1994). *Wavelets for kids*. Available at `www.diku.dk/users/jda/biosignal/kidsA.pdf`