Estimating invertible functional time series

Alexander Aue and Johannes Klepsch

Abstract This contribution discusses the estimation of an invertible functional time series through fitting of functional moving average processes. The method uses a functional version of the innovations algorithm and dimension reduction onto a number of principal directions. Several methods are suggested to automate the procedures. Empirical evidence is presented in the form of simulations and an application to traffic data.

1 Introduction

Functional time series have come into the center of statistics research at the confluence of functional data analysis and time series analysis. Some of the more and most recent contributions in this area include Aston and Kirch [1, 2] and Aue et al. [7] who dealt with the detection and estimation of structural breaks in functional time series, Chakraborty and Panaretos [9] who covered functional registration and related it to optimal transport problems, Horváth et al. [13] and Aue and van Delft [8] who developed stationarity tests in the time and frequency domain, respectively, Hörmann et al. [12] who introduced methodology for the detection of periodicities, Hörmann et al. [11] and Aue et al. [5] who proposed models for heteroskedastic functional time series, van Delft and Eichler [20] who defined a framework for locally stationary time series, Kowal et al. [16] who developed Bayesian methodology for a functional dynamic linear model, Raña et al. [18, 19] who discussed outlier detection in functional time series and provided methodology for the construction of

Alexander Aue
Department of Statistics, University of California, One Shields Avenue, Davis, CA 95616, USA, e-mail: aaue@ucdavis.edu

Johannes Klepsch
Center for Mathematical Sciences, Technische Universität München, 85748 Garching, Boltzmannstraße 3, Germany, e-mail: j.klepsch@tum.de

bootstrap confidence intervals for nonparametric regression under dependence, respectively, and Paparoditis [17] who introduced a sieve bootstrap procedure.

A general framework for functional time series allowing for elegant derivations of large-sample results was put forward in Hörmann and Kokoszka [10]. This paper introduced a concept to measure closeness of functional time series to certain functional moving average processes. It was then exploited that the latter have non-trivial autocovariance operators only for finitely many lags in order to derive large-sample results concerning the validity of functional principal components analysis in a dependent setting and change-point analysis (see [4]), among others. The focus of Aue and Klepsch [6] was not on theoretical properties but on the more practical question of how to estimate an invertible functional time series. This was achieved by functional moving average model fitting. The fitting process involved an application of the functional innovations algorithm, whose population properties were derived in Klepsch and Klüppelberg [14]. This algorithm can be used to estimate the operators in the causal representation of a functional time series. The consistency of these estimates is the main result. For practical purposes, the proposed method requires the selection of the dimension reduction space through both model selection and testing approaches. Several methods are proposed and then evaluated in a simulation study and in an application to vehicle traffic data.

The remainder is organized as follows. Section 2 introduces the setting and the method for estimating functional moving average processes. Several algorithms for practical implementation are discussed in Section 3. Section 4 gives a glimpse on large-sample theory. Section 5 briefly covers empirical aspects.

2 Estimation methodology

Let $H = L^2[0,1]$ be the Hilbert space of square-integrable functions on $[0,1]$ equipped with the standard norm $\|\cdot\|$ defined by the inner product $\langle x,y\rangle = \int_0^1 x(s)y(s)ds$, for $x,y \in H$. Let $(\Omega,\mathscr{A},P)$ be a probability space and denote by $L_H^2 = L^2(\Omega,\mathscr{A},P)$ the space of square integrable random functions taking values in H, noting that L_H^2 is a Hilbert space with inner product $E[\langle X,Y\rangle]$, for $X,Y \in L_H^2$. A functional linear process with values in L_H^2 is given by the series expansion

$$X_j = \sum_{\ell=0}^{\infty} \psi_\ell \varepsilon_{j-\ell}, \qquad j \in \mathbb{Z}, \tag{1}$$

where $(\psi_\ell\colon \ell \in \mathbb{N}_0)$ is a sequence in the space of bounded linear operators acting on H and $(\varepsilon_j\colon j \in \mathbb{Z})$ a sequence of independent, identically distributed random functions in L_H^2. A functional time series $(X_j\colon j \in \mathbb{Z})$ is invertible if it admits the series expansion

$$X_j = \sum_{\ell=1}^{\infty} \pi_\ell X_{j-\ell} + \varepsilon_j, \qquad j \in \mathbb{Z}, \tag{2}$$

where $(\pi_\ell\colon \ell \in \mathbb{N})$ is a sequence of bounded linear operators such that $\sum_{\ell=1}^\infty \|\pi_\ell\|_{\mathscr{L}} < \infty$, with $\|A\|_{\mathscr{L}} = \sup_{\|x\|\le 1}\|Ax\|$ for A bounded and linear.

If $X \in L_H^2$ with $E[X]=0$, then its covariance operator exists and admits a spectral representation; that is,

$$C_X(y) = E[\langle X, y\rangle X] = \sum_{i=1}^\infty \lambda_i \langle y, \nu_i\rangle \nu_i, \qquad y \in H, \tag{3}$$

where $(\lambda_i\colon i\in\mathbb{N})$ and $(\nu_i\colon i \in \mathbb{N})$ denote the eigenvalues and eigenfunctions, respectively. If $X,Y \in L_H^2$ with $E[X],E[Y]=0$, the cross covariance operator exists and is given by

$$C_{XY}(y) = E[\langle X, y\rangle Y], \qquad y\in H. \tag{4}$$

Introducing $x\otimes y(\cdot) = \langle x,\cdot\rangle y$ for $x,y\in H$, the lag-h autocovariance operator of a stationary functional time series $(X_j\colon j\in\mathbb{Z})$ may be written as $C_{X;h} = E[X_0\otimes X_h]$, for $h \in \mathbb{Z}$.

Let $\mathscr{V}_d = \overline{\mathrm{sp}}\{\nu_1,\dots,\nu_d\}$ be the subspace generated by the first d principal directions and let $P_{\mathscr{V}_d}$ be the projection operator from H to $\mathscr{V}_d$. For an increasing sequence $(d_i\colon i\in\mathbb{N})\subset\mathbb{N}$ define $X_{d,j} = P_{\mathscr{V}_{d_i}} X_j$, $j\in\mathbb{Z}$, $i\in\mathbb{N}$ and denote by $\tilde{\mathscr{F}}_{n,k}$ the smallest subspace containing $X_{d_k,n},\dots,X_{d_1,n-k}$ that is closed with respect to bounded, linear operators. The best linear predictor of X_{n+1} given $\tilde{\mathscr{F}}_{n,k}$ is then

$$\tilde{X}_{n+1,k} = P_{\tilde{\mathscr{F}}_{n,k}}(X_{n+1}) = \sum_{i=1}^k \theta_{k,i}(X_{d_{k+1-i},n+1-i} - \tilde{X}_{n+1-i,k-i}), \tag{5}$$

where $\tilde{X}_{n-k,0}=0$. This is the form required for the innovations algorithm developed in Klepsch and Klüppelberg [14] that provides a solution to recursively compute the coefficients $\theta_{k,1},\dots,\theta_{k,k}$ on the population level. However, for practical purposes, an estimated version of (5) is needed. Define

$$\hat{\mathscr{V}}_{d_i} = \overline{\mathrm{sp}}\{\hat{\nu}_1,\dots,\hat{\nu}_{d_i}\} \qquad \text{and} \qquad \hat{P}_{(k)} = \mathrm{diag}(P_{\hat{\mathscr{V}}_{d_k}},\dots,P_{\hat{\mathscr{V}}_{d_1}}),$$

where $\hat{\nu}_i$ denotes the ith eigenfunction of the sample covariance operator $\hat{C}_X$. The best linear predictor can now be computed with the functional innovations algorithm given in Section 3. The large-sample behavior is presented in Section 4.

3 Algorithms

The following algorithm details how an estimate of the best linear predictor may be computed by recursion. Denote by A^* the adjoint of an operator A.

Functional innovations algorithm. The best linear predictor $\tilde{X}_{n+1,k}$ of X_{n+1} given $\tilde{\mathscr{F}}_{n,k}$ can be computed by the recursions

$$\begin{aligned}
\tilde{X}_{n-k,0} &= 0 \qquad \text{and} \qquad \hat{V}_1 = P_{\hat{\mathscr{V}}_{d_1}} \hat{C}_X P_{\hat{\mathscr{V}}_{d_1}}, \\
\tilde{X}_{n+1,k} &= \sum_{i=1}^{k} \hat{\theta}_{k,i}(X_{d_{k+1-i},n+1-i} - \tilde{X}^{k-i}_{n+1-i}), \\
\hat{\theta}_{k,k-i} &= \Big(P_{\hat{\mathscr{V}}_{d_{k+1}}} \hat{C}_{X;k-i} P_{\hat{\mathscr{V}}_{d_{i+1}}} - \sum_{j=0}^{i-1} \hat{\theta}_{k,k-j} \hat{V}_j \hat{\theta}^*_{i,i-j} \Big) \hat{V}_i^{-1}, \qquad i = 1, \ldots, k-1, \\
\hat{V}_k &= \hat{C}_{X_{d_{k+1}} - \tilde{X}_{n+1,k}} = \hat{C}_{X_{d_{k+1}}} - \sum_{i=0}^{k-1} \hat{\theta}_{k,k-i} \hat{V}_i \hat{\theta}^*_{k,k-i}.
\end{aligned} \tag{6}$$

Application of the algorithm requires the selection of the d_i and and also the FMA order q. The selection of the former can be achieved through the following portmanteau test for independence. Here all d_i are set to the same value.

Determining the principal subspace by testing for independence.

(1) Given functions $X_1, \ldots, X_n$, estimate $\hat{\lambda}_1, \ldots, \hat{\lambda}_n$ and $\hat{\nu}_1, \ldots, \hat{\nu}_n$. Select d^* such that

$$\mathrm{TVE}(d^*) = \frac{\sum_{i=1}^{d^*} \hat{\lambda}_i}{\sum_{i=1}^{n} \hat{\lambda}_i} \geq P$$

for some prespecified $P \in (0,1)$.

(2) Let $f_h(\ell,\ell')$ and $b_h(\ell,\ell')$ denote the (ℓ,ℓ')th entries of $C^{-1}_{\mathbf{X}^*;0} C_{\mathbf{X}^*;h}$ and $C_{\mathbf{X}^*;h} C^{-1}_{\mathbf{X}^*;0}$, respectively, and $(\mathbf{X}^*_j \colon j \in \mathbb{Z})$ the process consisting of the $d+1$st to $d+l$th principal directions of C_X. If

$$Q_n^{d^*} = n \sum_{h=1}^{\bar{h}} \sum_{\ell,\ell'=d^*+1}^{d^*+l} f_h(\ell,\ell') b_h(\ell,\ell') > q_{\chi^2_{d^{*2}\bar{h}}},$$

set $d^* = d^* + 1$.

(3) If $Q_n^{d^*} \leq q_{\chi^2_{d^{*2}\bar{h}}}$, apply the functional innovations algorithm with $d_i = d^*$.

Once d is selected, the order of the resulting VMA process can be determined with a Ljung–Box test or an AICC criterion. Both are described next.

(I) Order selection with Ljung–Box test.

(1) Test the null hypothesis $H_0 \colon C_{\mathbf{X};h} = 0$ for all $h \in [\underline{h}, \overline{h}]$ with the test statistic

$$Q_{\underline{h},\bar{h}} = n^2 \sum_{h=\underline{h}}^{\bar{h}} \frac{1}{n-h} \mathrm{tr}\big(\hat{C}^{\top}_{\mathbf{X};h} \hat{C}^{-1}_{\mathbf{X};0} \hat{C}_{\mathbf{X};h} \hat{C}^{-1}_{\mathbf{X};0} \big),$$

which is asymptotically $\chi^2_{d^2(\bar{h}-\underline{h}-1)}$-distributed.

(2) Iteratively compute $Q_{1,\bar{h}}, Q_{2,\bar{h}}, \ldots$ and select q as the largest $\underline{h}$ such that $Q_{\underline{h},\bar{h}}$ is significant but $Q_{\underline{h}+h,\bar{h}}$ is insignificant for all h.

(II) Order selection with AICC criterion.

(1) Choose the order of the FMA process as the minimizer of

$$\mathrm{AICC}(q) = -2\ln L(\Theta_1,\ldots,\Theta_q,\Sigma) + \frac{2nd(qd^2+1)}{nd-qd^2-2}$$

where $\Theta_1,\ldots,\Theta_1$ are the VMA matrices and Σ the covariance matrix.

Another option is provided by the following FPE-type criterion that selects d and q jointly. Is is an adaptation of a similar criterion for FAR processes put forward in Aue et al. [3].

Determination of principal subspace and order selection with FPE criterion.

(1) Select (d,q) as minimizer of

$$\mathrm{fFPE}(d,q) = \frac{n+qd}{n}\mathrm{tr}(\hat{\mathbf{V}}_n) + \sum_{i>d}\hat{\lambda}_i,$$

where $\hat{\mathbf{V}}_n$ is the matrix version of $\hat{V}_n$ in (6).

4 Large-sample properties

The following theorem states that $\hat{\Gamma}_k$ is a consistent estimator for Γ_k.

Theorem 1. *If $(X_j\colon j\in\mathbb{Z})$ defined in (1) is such that $\sum_{m=1}^{\infty}\sum_{\ell=m}^{\infty}\|\psi_\ell\|_{\mathscr{L}}<\infty$ and $E[\|\varepsilon_0\|^4]<\infty$, then $(n-k)E[\|\hat{\Gamma}_k-\Gamma_k\|_{\mathscr{N}}^2]\le kU_X$, where $\|\cdot\|_{\mathscr{N}}$ denotes nuclear norm and U_X a constant that does not depend on n.*

To discuss the consistency of the estimators in the causal and invertible representations, further conditions are needed. As $n\to\infty$, let $k=k_n\to\infty$ and $d_k\to\infty$ such that

$$k^{1/2}(n-k)^{-1/2}\alpha_{d_k}^{-2}\to 0,$$

$$k^{1/2}\alpha_{d_k}^{-1}\left(\sum_{\ell>k}\|\pi_\ell\|_{\mathscr{L}} + \sum_{\ell=1}^{k}\|\pi_\ell\|_{\mathscr{L}}\sum_{i>d_{k+1-\ell}}\lambda_i\right)\to 0,$$

$$k^{3/2}\alpha_{d_k}^{-2}n^{-1}\left(\sum_{\ell=1}^{d_k}\delta_\ell^{-2}\right)^{1/2}\to 0, \qquad (7)$$

where α_{d_k} is related to the spectral gaps of C_X and α_{d_n} is the infimum of the eigenvalues of the spectral density operator of $((\langle X_n,\nu_1\rangle,\ldots,\langle X_n,\nu_{d_n}\rangle)^T : n\in\mathbb{N})$.

Theorem 2. *Under the assumptions of Theorem 1 and the above conditions, for all $x\in H$ and $i\in\mathbb{N}$ as $n\to\infty$, $\|(\hat{\beta}_{k,i}-\pi_i)(x)\|\xrightarrow{p}0$, and $\|(\hat{\theta}_{k,i}-\psi_i)(x)\|\xrightarrow{p}0$. If the*

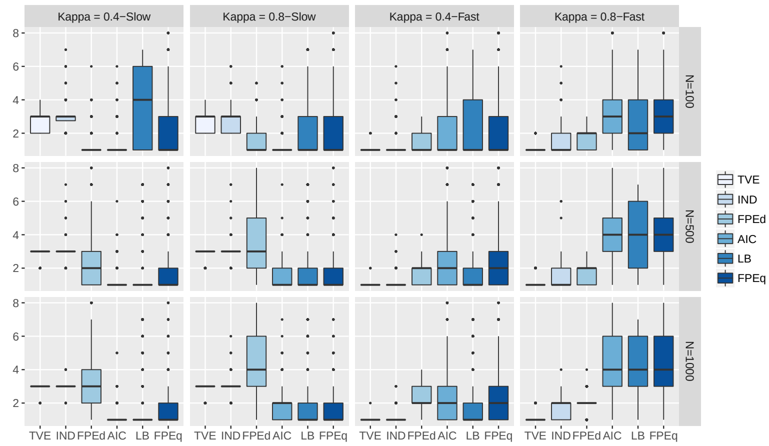

Fig. 1 Model selection for different FAR(1) processes. The left three plots in each small figure give the d chosen by total variation explained with $P = 0.8$ (TVE), the test for independence (IND) and the functional FPE criterion (FPEd). The right three plots in each small figure give the selected order q by AICC, Ljung–Box and fFPE.

operators $(\psi_\ell : \ell \in \mathbb{N})$ *and* $(\pi_\ell : \ell \in \mathbb{N})$ *are Hilbert–Schmidt, then the convergence is uniform.*

Detailed proofs of both theorems may be found in Aue and Klepsch [6].

5 Empirical results

As an illustration of the proposed fitting method, a simulation is provided in which an FAR(1) process is approximated through an FMA(q) process, where the process is generated as outlined in Aue et al. [3], choosing various norms κ of the FAR operator and fast and slow decays of eigenvalues of the covariance operator. Model selection results for the different methods are provided in Figure 1, noting that FMA models are fit to an FAR time series.

Figure 2 displays 1440 curves of average velocity per minute obtained at a fixed measurement station on A92 Autobahn in southern Germany. Klepsch et al. [15] indicate that this functional time series is stationary and that an FMA fit may be appropriate. Applying the functional innovations algorithm together with any of the proposed procedures to select d and q leads to a first-order dynamics. Additional information is provided in Aue and Klepsch [6].

Acknowledgements This research was supported by NSF grants DMS 1305858 and DMS 1407530. The velocity data was provided by Autobahndirektion Südbayern.

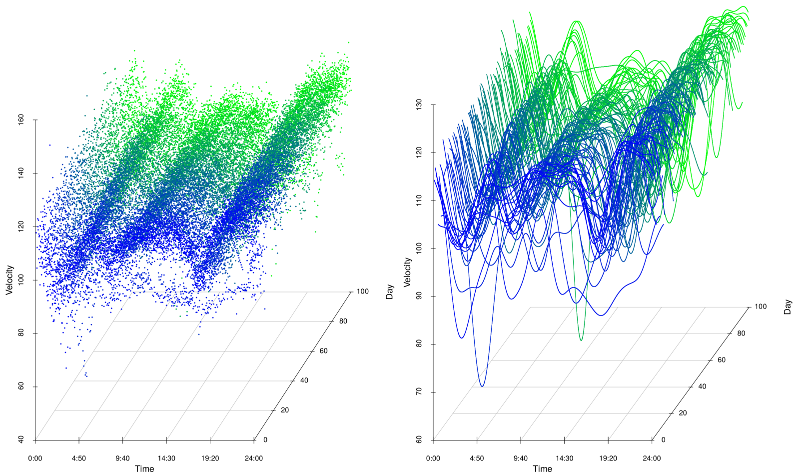

Fig. 2 Discrete velocity observations (left) and corresponding velocity functions (right).

References

1. Aston, J.A.D., Kirch, C.: Detecting and estimating changes in dependent functional data. J. Mult. Anal. **109**, 204–220 (2012).
2. Aston, J.A.D., Kirch, C.: Evaluating stationarity via change-point alternatives with applications to FMRI data. Ann. Appl. Statist. **6**, 1906–1948 (2012).
3. Aue, A., Dubart Norinho, D., Hörmann, S.: On the prediction of stationary functional time series. J. Am. Stat. Assoc. **110**, 378–392 (2015)
4. Aue, A., Horváth, L.: Structural breaks in time series. J. Time Ser. Anal. **34**, 1–16 (2013).
5. Aue, A., Horváth, L., Pellatt, D.: Functional generalized autoregressive conditional heteroskedasticity. J. Time Ser. Anal. **38**, 3–21 (2017).
6. Aue, A., Klepsch, J.: Estimating functional time series by moving average model fitting. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1701.00770.
7. Aue, A., Rice, G., Sönmez, O.: Detecting and dating structural breaks in functional data without dimension reduction. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1511.04020.
8. Aue, A., van Delft, A.: Testing for stationarity of functional time series in the frequency domain. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1701.01741.
9. Chakraborty, A., Panaretos, V.M.: Functional registration and local variations. ArXiv preprint (2017). Available online at https://arxiv.org/abs/1702.03556
10. Hörmann, S., Kokoszka, P.: Weakly dependent functional data. Ann. Statist. **38**, 1845–1884 (2010).
11. Hörmann, S., Horváth, L., Reeder, R.: A functional version of the ARCH model. Econometr. Th. **29**, 267–288 (2013).
12. Hörmann, S., Kokoszka, P., Nisol, G.: Detection of periodicity in functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1607.02017.
13. Horváth, L., Kokoszka, P., Rice, G.: Testing stationarity of functional time series. J. Econ. **179**, 66–82 (2014).

14. Klepsch, J., Klüppelberg, C.: An innovations algorithm for the prediction of functional linear processes. J. Mult. Anal. **155**, 252–271 (2017).
15. Klepsch, J., Klüppelberg, C., Wei, T.: Prediction of functional ARMA processes with an application to traffic data. Econ. Statist. **1**, 128–149 (2017).
16. Kowal, D.R, Matteson, D.S., Ruppert, D.: A Bayesian multivariate functional dynamic linear model. J. Am. Statist. Assoc. (2016) doi: 10.1080/01621459.2016.1165104
17. Paparoditis, E.: Sieve bootstrap for functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1609.06029.
18. Raña, P., Aneiros, G., Vilar, J.M.: Detection of outliers in functional time series. Environm. **26**, 178–191 (2015).
19. Raña, P., Aneiros, G., Vilar, J.M., Vieu, P.: Bootstrap confidence intervals for nonparametric regression under dependence. Electr. J. Statist. **10**, 1973–1999 (2016).
20. Van Delft, A., Eichler, M.: Locally stationary functional time series. ArXiv preprint (2016). Available online at https://arxiv.org/abs/1602.05125.