Supplemental Material: Predictive inference for locally stationary time series with an application to climate data

Srinjoy Das
Department of Electrical
and Computer Engineering
Univ. of California—San Diego
La Jolla, CA 92093, USA
email: s2das@ucsd.edu

Dimitris N. Politis
Department of Mathematics
Univ. of California—San Diego
La Jolla, CA 92093-0112, USA
email: dpolitis@ucsd.edu

A Appendix: Basic Model-Free and Limit Model-Free Bootstrap Algorithms

This section describes in detail algorithms A.1 and A.2 for the construction of Model-Free and Limit Model-Free algorithms as described in (Politis, 2015).

Define the *predictive root* to be the error in prediction, i.e.,

$$Y_{n+1} - \Pi(\hat{g}_{n+1}, \underline{Y}_n, \hat{F}_n) \tag{1}$$

where $\Pi(\hat{g}_{n+1}, \underline{Y}_n, \hat{F}_n)$ is our chosen point predictor of Y_{n+1}, and $\hat{g}_{n+1}$ is our estimate of function g_{n+1} based on the data $\underline{Y}_n$.

Given bootstrap data $\underline{Y}^*_n$ and Y^*_{n+1}, the bootstrap predictive root is the error in prediction in the bootstrap world, i.e.,

$$Y^*_{n+1} - \Pi(\hat{g}^*_{n+1}, \underline{Y}_n, \hat{F}_n) \tag{2}$$

where $\hat{g}^*_{n+1}$ is our estimate of function g_{n+1} based on the bootstrap data $\underline{Y}^*_n$.

Remark A.1 Note that eq. (2) depends on the bootstrap data $\underline{Y}^*_n$ only through the estimated function $\hat{g}^*_{n+1}$; both the predictor $\Pi(\hat{g}^*_{n+1}, \underline{Y}_n, \hat{F}_n)$ and the construction of future value Y^*_{n+1} in the sequel are based on the true dataset $\underline{Y}_n$ in order to give validity to the prediction intervals *conditionally* on the data $\underline{Y}_n$.

Algorithm A.1 Model-free bootstrap for prediction intervals for Y_{n+1}

1. *Based on the data $\underline{Y}_n$, estimate the transformation H_n and its inverse H_n^{-1} by $\hat{H}_n$ and $\hat{H}_n^{-1}$ respectively. In addition, estimate g_{n+1} by $\hat{g}_{n+1}$.*

2. *Use $\hat{H}_n$ to obtain the transformed data, i.e., $(\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)})' = \hat{H}_n(\underline{Y}_n)$. By construction, the variables $\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)}$ are approximately i.i.d.; let $\hat{F}_n$ denote their empirical distribution.*

   (a) *Sample randomly (with replacement) the data $\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)}$ to create the bootstrap pseudo-data $\varepsilon_1^*, ..., \varepsilon_n^*$.*

(b) *Use the inverse transformation* $\hat{H}_n^{-1}$ *to create pseudo-data in the* Y *domain, i.e., let* $\underline{Y}_n^* = (Y_1^*, ..., Y_n^*)' = \hat{H}_n^{-1}(\varepsilon_1^*, ..., \varepsilon_n^*)$.

(c) *Calculate a bootstrap pseudo-response* Y_{n+1}^* *as the point* $\hat{g}_{n+1}(\underline{Y}_n, \varepsilon)$ *where* ε *is drawn randomly from the set* $(\varepsilon_1^{(n)}, ..., \varepsilon_n^{(n)})$.

(d) *Based on the pseudo-data* $\underline{Y}_n^*$, *estimate the function* g_{n+1} *by* $\hat{g}_{n+1}^*$ *respectively.*

(e) *Calculate a bootstrap root replicate using eq. (2).*

3. *Steps (a)—(e) in the above should be repeated a large number of times (say* B *times), and the* B *bootstrap root replicates should be collected in the form of an empirical distribution whose* α*—quantile is denoted by* $q(\alpha)$.

4. *A* $(1-\alpha)100\%$ *equal-tailed prediction interval for* Y_{n+1} *is given by*

$$[\Pi + q(\alpha/2),\ \Pi + q(1-\alpha/2)] \tag{3}$$

where Π *is short-hand for* $\Pi(\hat{g}_{n+1}, \underline{Y}_n, \hat{F}_n)$.

Sometimes, the empirical distribution $\hat{F}_n$ converges to a limit distribution F that is of known form (perhaps after estimating a finite-dimensional parameter). Using it instead of the empirical $\hat{F}_n$ results into the Limit Model-Free (LMF) resampling algorithm that is given below. Note that now the point predictor Π is no more a function of $\hat{F}_n$ but of F. Hence, the LMF predictive root is denoted by

$$Y_{n+1} - \Pi(\hat{g}_{n+1}, \underline{Y}_n, F) \tag{4}$$

whose distribution can be approximated by that of the LMF bootstrap predictive root

$$Y_{n+1}^* - \Pi(\hat{g}_{n+1}^*, \underline{Y}_n, F). \tag{5}$$

Algorithm A.2 Limit Model-free (LMF) bootstrap for prediction intervals for Y_{n+1}

1. *Based on the data $\underline{Y}_n$, estimate the transformation H_n and its inverse H_n^{-1} by $\hat{H}_n$ and $\hat{H}_n^{-1}$ respectively. In addition, estimate g_{n+1} by $\hat{g}_{n+1}$.*

2. *(a) Generate bootstrap pseudo-data $\varepsilon_1^*, ..., \varepsilon_n^*$ in an i.i.d. manner from F.*

   (b) Use the inverse transformation $\hat{H}_n^{-1}$ to create pseudo-data in the Y domain, i.e., let $\underline{Y}_n^* = (Y_1^*, ..., Y_n^*)' = \hat{H}_n^{-1}(\varepsilon_1^*, ..., \varepsilon_n^*)$.

   (c) Calculate a bootstrap pseudo-response Y_{n+1}^* as the point $\hat{g}_{n+1}(\underline{Y}_n, \varepsilon)$ where ε is a random draw from distribution F.

   (d) Based on the pseudo-data $\underline{Y}_n^*$, estimate the function g_{n+1} by $\hat{g}_{n+1}^*$ respectively.

   (e) Calculate a bootstrap root replicate using eq. (5).

3. *Steps (a)—(e) in the above should be repeated a large number of times (say B times), and the B bootstrap root replicates should be collected in the form of an empirical distribution whose α—quantile is denoted by $q(\alpha)$.*

4. *A $(1-\alpha)100\%$ equal-tailed prediction interval for Y_{n+1} is given by*

$$[\Pi + q(\alpha/2),\ \Pi + q(1-\alpha/2)] \tag{6}$$

   where Π is short-hand for $\Pi(\hat{g}_{n+1}, \underline{Y}_n, F)$.

B Appendix: RAMPFIT algorithm for analyzing climate data with transitions

The RAMPFIT algorithm which can handle uneven time-spacing in observations was proposed by (Mudelsee, 2000) for performing regression on climate data which shows transitions such as the speleothem dataset considered in this paper. However RAMPFIT was not originally designed to handle arbitrary local stationarity which may be present in data.

Here we briefly outline the steps in RAMPFIT used to obtain point prediction estimates which are used for comparison with their Model-Based and Model-Free counterparts.

Define $x(i) = X(t(i))$ where $(X_t, t \in \mathbf{R})$ is an underlying continuous-time stochastic process. For a time series $x(i)$ measured at times $t(i), i = 1, \ldots, n$, the model under consideration is (Mudelsee, 2000):

$$x(i) = x_{fit}(i) + \epsilon(i) \tag{7}$$

It is assumed that the errors $\epsilon(i)$ are heteroskedastic and are distributed as $N(0, \sigma(i)^2)$. The fitted model is a ramp function as defined below:

$$x_{fit}(t) = \begin{cases} x1, & for\ t \leq t1, \\ x1 + (t - t1)(x2 - x1)/(t2 - t1), & for\ t1 \leq t \leq t2, \\ x2, & for\ t \geq t2 \end{cases} \tag{8}$$

Here $t1$ and $t2$ denote the start and end of the ramp and $x1$, $x2$ denote the corresponding values at those points. The regression model is fitted to data $\{t(i),\, x(i)\}_{i=1}^{n}$ by minimizing the weighted sum of squares as given below:

$$SSQW(t1, x1, t2, x2) = \sum_{i=1}^{n} \frac{[x(i) - x_{fit}(i)]^2}{\sigma(i)^2} \tag{9}$$

Owing to the non-differentiabilities at $t1$ and $t2$, RAMPFIT does a search over a range of values supplied for these 2 values and chooses the values $(\hat{t}1, \hat{x}1, \hat{t}2, \hat{x}2)$ for which the $SSQW$ is minimum. In addition since $\sigma(i)$ is not known an initial guess of this is supplied to the algorithm following which the $\sigma(i)$ values are recalculated from the obtained residuals. The estimates $(\hat{t}1, \hat{x}1, \hat{t}2, \hat{x}2)$ are then regenerated. These steps are repeated till MSE values of point prediction converge. The full algorithm is described below:

Algorithm B.1 *RAMPFIT REGRESSION*

1. *Set initial estimate of* $\sigma(i) = i$ *with* $i = 1, \ldots, n$
2. *Set search ranges* $[t1_{min}, t1_{max}]$ *and* $[t2_{min}, t2_{max}]$ *for values of* $t1$ *and* $t2$
3. *Calculate SSQW using (8) and (9) over this grid of* $t1$ *and* $t2$ *values; denote a typical point in this grid as* $(\bar{t}1, \bar{t}2)$
4. *Determine* $(\hat{t}1, \hat{x}1, \hat{t}2, \hat{x}2)$ = *argmin* $[SSQW(\bar{t}1, \hat{x}1, \bar{t}2, \hat{x}2)]$ *and obtain* x_{fit}
5. *Calculate residuals* $e(i) = x(t(i)) - x_{fit}(t(i))$
6. *Re-estimate the variance* $\sigma(i)$ *from* $e(i)$ *using k-nearest-neighbour smoothing*
7. *Repeat steps (2) to (6) above till MSE values converge.*

C Appendix: Diagnostics for Model-Free Inference

The steps outlined in Section 3.1 for Model-Free inference describe a transformation of the data first to uniform, and then to standard normal distributions. Since the transformation involves quantities that need to be estimated, some diagnostics are suggested to ensure that the practical, data-based transformation gives the desired results.

C.1 QQ-plots after uniformization

The success of the uniformization step outlined in Section 3.1 can be visually verified using QQ-plots of the obtained uniform samples versus samples obtained from an ideal uniform distribution which is available in standard statistical software such as R. Any deviations in these curves from linearity should be closely investigated for possible issues wrt choice of bandwidth during cross-validation as it can impact both point prediction and prediction interval generation.

C.2 Shapiro-Wilk test for joint normality

As discussed before a convenient way to ensure both the smoothness and data-based consistent estimation of $\mathcal{L}(Y_t, Y_{t-1}, \ldots, Y_{t-m+1})$ for any m is to assume that, for all t,

$$Y_t = \mathbf{f}_t(W_t) \tag{10}$$

for some function $\mathbf{f_t}(w)$ that is smooth in both arguments t and w, and some strictly stationary and weakly dependent, univariate time series W_t; without loss of generality, we will assume that W_t is a Gaussian time series. In the Model-free setup, assuming Eq. (10) cannot be seen as restrictive; indeed, Y_t can have an arbitrary marginal distribution—that is time-changing as well—and the underlying dependence can be strong (or not), as inherited by the W_t. Therefore, it should not be surprising that it is difficult to check whether a general condition such as (10) holds for a particular dataset at hand.

Note, however, that assumption (10) was invoked in order to ensure the joint normality of the elements of $\underline{Z}_n = (Z_1, \ldots, Z_n)'$ defined in Section 3.6; see Lemma 3.1. Hence, we may check $\underline{Z}_n$ for normality which would then serve as an implicit diagnostic as to whether Eq. (10) holds true or not.

Marginal normality of the Z_t can be verified by gauging linearity of QQ-plots versus the standard normal distribution. Furthermore, by the Cramer-Wold device, any linear combination of jointly normal variables is univariate normal; this can be used to empirically verify whether the joint normality requirement is violated by taking an arbitrary linear combination i.e. for example a pair or triplet of variables from the set $\underline{Z}_n = (Z_1, \ldots, Z_n)'$, and gauging its normality using the Shapiro-Wilk test. An example is provided in Figure 1 where, for a given λ, we form the linear combination $(1-\lambda)Z_t + \lambda Z_{t+1}$ for all t, and calculate the mean value of the associated Shapiro-Wilk test statistic; this can be done over a range of λ values. As can be seen from the plot, sufficiently high values of the test statistic are obtained indicating that we can not reject the hypothesis of joint normality. Further tests can be done by forming linear combinations over pairs of non-successive values of Z_t, i.e., checking the normality of $(1 - \lambda)Z_t + \lambda Z_{t+k}$ for some different values of k.

C.3 Kolomogorov-Smirnov test for i.i.d. standard normal samples

Provided that the inputs are jointly normal the whitening transformation described in Section 3.6 produces i.i.d. standard normal variables. The covariance matrix used in this step can be derived either by fitting a causal AR(p) model to $\underline{Z}_n = (Z_1, \ldots, Z_n)'$ or using the flat-top kernel banded, tapered estimator outlined in (McMurry & Politis, 2010). To verify that the data generated after whitening are standard normal a Kolmogorov-Smirnov test can be used with the reference distribution as $N[0, 1]$.

C.4 Independence test of standard normal samples

The success of the Model-Free procedure involves the ability to produce i.i.d. data after a series of invertible transformations. In the case of Locally Stationary Time Series independence of the data produced at the final step after applying the whitening transformation can be verified visually using an autocorrelation function (ACF) plot as the data are approximately standard normal. An example is given in Figure 2 where it can be noticed from the ACF plot that the Model-Free transformations were successful in producing decorrelated (and therefore i.i.d.) normal data.

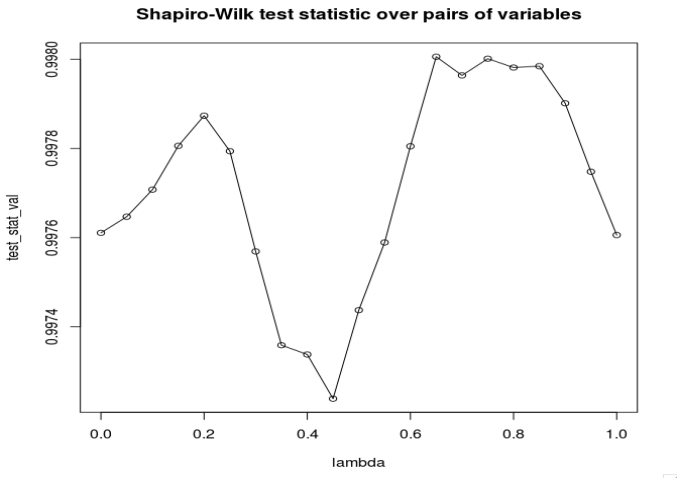

Figure 1: Values of Shapiro-Wilk test statistic for joint normality test. Note that corresponding p-values range from 0.09 to 0.29.

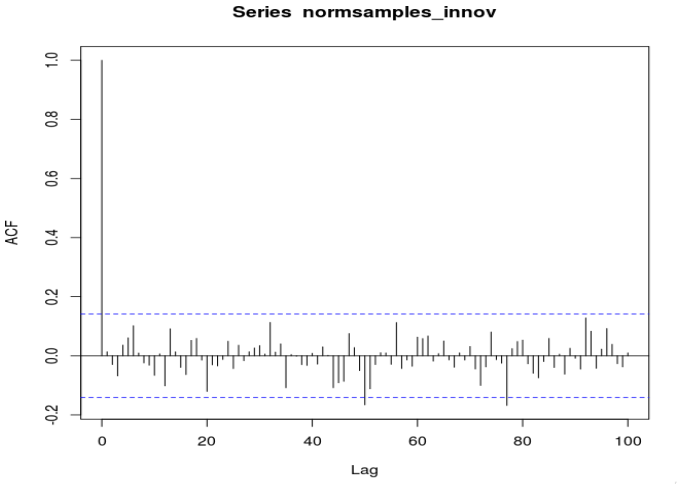

Figure 2: Autocorrelation plot showing decorrelation/independence of data after whitening transformation

References

McMurry, T. L., & Politis, D. N. (2010). Banded and tapered estimates for autocovariance matrices and the linear process bootstrap. Journal of Time Series Analysis, 31(6), 471–482.

Mudelsee, M. (2000). Ramp function regression: a tool for quantifying climate transitions. Computers & Geosciences, 26(3), 293–307.

Politis, D. N. (2015). Model-free prediction and regression. Springer, New York.