are enough to imply asymptotic validity of the bootstrap prediction interval. Also note that part (ii) of Definition 3.6.1 is the condition needed in order to show that the bootstrap can yield asymptotically valid *confidence intervals* for the conditional mean $\mu(\cdot)$. In many cases in the literature, this condition has been already established; we can build upon this for the purpose of constructing pertinent prediction intervals.

Consider again the heteroscedastic model (3.1). Much of the above discussion carries over *verbatim*; for example, the MSE–optimal predictor of Y_{n+1} given $X_{n+1} = x_{n+1}$ is still $\hat{Y}_{n+1} = \hat{\mu}(x_{n+1})$. The only difference is that the predictive root now is expressed as

$$Y_{n+1} - \hat{Y}_{n+1} = \sigma(x_{n+1})\varepsilon_{n+1} + A_\mu, \tag{3.25}$$

and the bootstrap predictive root as

$$Y^*_{n+1} - \hat{Y}^*_{n+1} = \hat{\sigma}(x_{n+1})\varepsilon^*_{n+1} + A^*_\mu \tag{3.26}$$

where $\hat{\sigma}(\cdot)$ is the (consistent) estimator of $\sigma(\cdot)$ that is employed in the bootstrap data generation mechanism. Hence, the following definition is immediate.

Definition 3.6.2 ***Asymptotic pertinence of bootstrap prediction intervals under heteroscedastic model*** (3.1). *Consider a bootstrap prediction interval for Y_{n+1} that is based on approximating the distribution of the predictive root $Y_{n+1} - \hat{Y}_{n+1}$ of eq.* (3.25) *by the distribution of the bootstrap predictive root $Y^*_{n+1} - \hat{Y}^*_{n+1}$ of eq.* (3.26). *The interval will be called* asymptotically pertinent *provided the bootstrap satisfies conditions (i)—(iii) or Definition 3.6.1 together with the additional requirement:*

(iv′) $\sigma(x_{n+1}) - \hat{\sigma}(x_{n+1}) \xrightarrow{P} 0.$

Furthermore, the bootstrap prediction interval for Y_{n+1} that is based on the approximating the distribution of the studentized *predictive root $(Y_{n+1} - \hat{Y}_{n+1})/\hat{V}_n$ by the distribution of the bootstrap studentized predictive root $(Y^*_{n+1} - \hat{Y}^*_{n+1})/\hat{V}^*_n$ will be called asymptotically pertinent if, in addition condition (iv) or Definition 3.6.1 also holds.*

Fact 3.6.1 *Under model* (3.1) *and standard regularity conditions, the model-based bootstrap prediction interval* (3.16) *will be asymptotically pertinent provided the bandwidth h is chosen in such a way that undersmoothing occurs, i.e., letting $h = o(n^{-1/5})$ when the kernel K is nonnegative. Otherwise, interval* (3.16) *will be asymptotically valid but not pertinent.*

Remark 3.6.2 Taking into account that $A_\mu = o_p(1)$ as $n \to \infty$, an immediate estimator for the (conditional) variance of the predictive root $Y_{n+1} - \hat{Y}_{n+1}$ under model (3.1) is simply $\hat{V}_n = \hat{\sigma}(x_{n+1})$. Therefore, condition (iv) or Definition 3.6.1 can be rewritten as $\hat{\sigma}(x_{n+1}) - \hat{\sigma}^*(x_{n+1}) \xrightarrow{P} 0$, i.e., it is just a bootstrap version of condition (iv′) or Definition 3.6.2. As a matter of fact, resampling in the heteroscedastic

estimator that yields a proper distribution function while maintaining its favorable asymptotic properties. The local linear versions of $\hat{D}_x(y)$ and $\bar{D}_x(y)$ using Hansen's (2004) adjustment are given by:

$$\hat{D}_x^{LL}(y) = \frac{\sum_{i=1}^n w_i^\diamond \mathbf{1}(Y_i \le y)}{\sum_{i=1}^n w_i^\diamond} \text{ and } \bar{D}_x^{LL}(y) = \frac{\sum_{i=1}^n w_i^\diamond \Lambda\left(\frac{y-Y_i}{h_0}\right)}{\sum_{i=1}^n w_i^\diamond}. \tag{4.7}$$

The weights $w_i^\diamond$ are defined by

$$w_i^\diamond = \begin{cases} 0 & \text{when } \hat{\beta}(x-X_i) > 1 \\ w_i(1-\hat{\beta}(x-X_i)) & \text{when } \hat{\beta}(x-X_i) \le 1 \end{cases} \tag{4.8}$$

where

$$w_i = \frac{1}{h} K\left(\frac{x-X_i}{h}\right) \text{ and } \hat{\beta} = \frac{\sum_{i=1}^n w_i(x-X_i)}{\sum_{i=1}^n w_i(x-X_i)^2}. \tag{4.9}$$

See Chap. 9 for an application of the above to a time series prediction problem.

Fact 4.2.1 *Under regularity conditions that include a well-behaved "density" $f(\cdot)$ (e.g., large-sample histogram) of the design points $x_1,\ldots,x_n$ and the assumption that, for all x, $D_x(y)$ is twice continuously differentiable as a function of y, it follows that $\bar{D}_x(y)$ satisfies an equation similar to Eq. (4.5), namely:*

$$\text{Var}\left(\bar{D}_x(y)\right) = O\left(\frac{1}{hn}\right) \text{ and } \text{Bias}\left(\bar{D}_x(y)\right) = O(h^2 + h_0^2) \tag{4.10}$$

assuming that $h_0 = o(h)$, $h \to 0$, $hn \to \infty$, and $\sqrt{hn}(h^3 + h_0^3) = o(1)$; see Theorem 6.2 of Li and Racine (2007). Furthermore, the two estimators $\bar{D}_x(y)$ and $\hat{D}_x(y)$ are asymptotically equivalent, *i.e., for any fixed x, $\sqrt{hn}\,(\bar{D}_x(y) - \hat{D}_x(y)) = o_p(1)$.*

Interestingly, although the two estimators $\bar{D}_x(y)$ and $\hat{D}_x(y)$ have Mean Squared Errors (MSE) that are of the same asymptotic order, smoothing may give a finite-sample advantage when the true $D_x(y)$ is smooth (at least twice continuously differentiable) as a function of y. Comparing Eq. (6.2) and (6.4) of Li and Racine (2007), it follows that:

$$\text{MSE}\left[\hat{D}_y(x)\right] - \text{MSE}\left[\bar{D}_y(x)\right] = c_{y,x}\frac{h_0}{nh} + o(\max\{h^4, \frac{1}{nh}\}) \tag{4.11}$$

where $c_{y,x} = C\frac{\partial}{\partial y} D_x(y)/f(x)$ for some constant $C \ge 0$.

Remark 4.2.4 (On choice of bandwidths) In order to minimize the asymptotic MSE of $\bar{D}_x(y)$, the optimal bandwidth specifications are $h \sim c_h n^{-1/5}$ and $h_0 \sim c_0 n^{-2/5}$ for some positive constants c_h, c_0. This suggests the following bandwidth choice rule-of-thumb which works reasonably well in practice: pick h via cross-validation, and then let $h_0 = h^2$.

NOTE: An improved method of local linear distribution estimation is given in: S. Das and D.N. Politis, `Nonparametric estimation of the conditional distribution at regression boundary points', The American Statistician, vol. 74, no. 3, pp. 232-242, 2020.

$$Y_t = \sum_{k=-\infty}^{\infty} \psi_k Z_{t-k} \tag{6.1}$$

where the coefficients ψ_k are (at least) square-summable, and the series $\{Z_t\}$ is i.i.d. with mean zero and variance $\sigma^2 > 0$. A linear time series $\{Y_t\}$ is called *causal* if $\psi_k = 0$ for $k < 0$, i.e., if

$$Y_t = \sum_{k=0}^{\infty} \psi_k Z_{t-k}. \tag{6.2}$$

Remark 6.1.1 Eq. (6.2) should not be confused with the Wold decomposition that *all* purely nondeterministic, stationary time series possess—see e.g. Brockwell and Davis (1991). In the Wold decomposition, the innovations $\{Z_t\}$ are only assumed to be a white noise and not i.i.d.; the i.i.d. assumption is of course much stronger.

A linear time series is called *invertible* if one can use eq. (6.1) to solve for Z_t in terms of present and past Y_ts in which case we can write

$$Y_t = \sum_{k=1}^{\infty} \phi_k Y_{t-k} + Z_t; \tag{6.3}$$

a typical assumption here is that the sequence ϕ_k is absolutely summable. For causal time series, invertibility occurs when the power series $\psi(s) = \sum_{k=0}^{\infty} \psi_k s^k$ has no roots on the unit disc. Similarly, for a time series satisfying eq. (6.3), causality occurs if the function $\phi(s) = \sum_{k=0}^{\infty} \phi_k s^k$ has no roots on the unit disc. Now it is not difficult to see that for a linear time series satisfying eq. (6.2) and (6.3) we have

$$E(Y_{n+1}|Y_n, Y_{n-1}, \ldots) = \sum_{k=1}^{\infty} \phi_k Y_{t-k}$$

where $E(Y_{n+1}|Y_n, Y_{n-1}, \ldots)$ denotes the conditional expectation given the infinite history. Hence, given the infinite past, the property of the optimal predictor being linear is shared by the class of linear time series that are *causal and invertible*.[1]

Under standard weak dependence conditions, it holds that

$$E(Y_0|Y_{-1}, Y_{-2}, \ldots, Y_{-m}) \to E(Y_0|Y_{-1}, Y_{-2}, \ldots) \text{ as } m \to \infty$$

for almost all sample paths of $\{Y_t,\ t < 0\}$. Using the assumed stationarity of $\{Y_t\}$ we can then write

$$E(Y_{n+1}|Y_n, Y_{n-1}, \ldots, Y_1) \simeq E(Y_{n+1}|Y_n, Y_{n-1}, \ldots) \tag{6.4}$$

for large n, i.e.,

$$E(Y_{n+1}|Y_n, Y_{n-1}, \ldots, Y_1) \simeq \sum_{k=1}^{n} \phi_k Y_{t-k}. \tag{6.5}$$

replace t by n+1

[1] A slight generalization of this statement is possible, i.e., replacing the i.i.d. assumption for $\{Z_t\}$ with a martingale difference assumption; see e.g. Politis (2009), or Kokoszka and Politis (2011).

Definition 7.2.1 ***Asymptotic pertinence of bootstrap prediction intervals under model*** (7.1). *Consider a bootstrap prediction interval for Y_{n+1} that is based on approximating the distribution of the predictive root $Y_{n+1} - \hat{Y}_{n+1}$ of eq.* (7.8) *by the distribution of the bootstrap predictive root $Y^*_{n+1} - \hat{Y}^*_{n+1}$ of eq.* (7.9). *The interval will be called* asymptotically pertinent *provided the bootstrap satisfies the following three conditions as $n \to \infty$ conditionally on $Y_{n-p+1} = y_{n-p+1}, \cdots, Y_n = y_n$.*

(i) $\sup_a |P(\varepsilon_{n+1} \le a) - P^*(\varepsilon^*_{n+1} \le a)| \xrightarrow{P} 0$, *presupposing that the error distribution is continuous.*

(ii) $|P(a_n A_\mu \le a) - P^*(a_n A^*_\mu \le a)| \xrightarrow{P} 0$ *for some sequence $a_n \to \infty$, and for all points a where the assumed nontrivial limit of $P(a_n A_\mu \le a)$ is continuous.*

(iii) ε^*_{n+1} and A^*_μ are independent in the bootstrap world—as their analogs are in the real world due to the causality assumption (7.3).

Furthermore, the bootstrap prediction interval for Y_{n+1} that is based on the approximating the distribution of the studentized *predictive root $(Y_{n+1} - \hat{Y}_{n+1})/\hat{V}_n$ by the distribution of the bootstrap studentized predictive root $(Y^*_{n+1} - \hat{Y}^*_{n+1})/\hat{V}^*_n$ will be called asymptotically pertinent if, in addition to (i)—(iii) above, the following also holds:*

(iv) $\hat{V}_n / \hat{V}^*_n \xrightarrow{P} 1$.

Consider now the heteroscedastic model (7.2). Much of the above discussion carries over *verbatim*; for example, our predictor of Y_{n+1} given $Y_1 = y_1, \ldots, Y_n = y_n$ is still $\hat{Y}_{n+1} = \hat{m}(y_n, \ldots, y_{n-p+1})$. The only difference is that the predictive root now is

$$Y_{n+1} - \hat{Y}_{n+1} = \sigma(y_n, \ldots, y_{n-p+1})\varepsilon_{n+1} + A_\mu, \tag{7.10}$$

and the bootstrap predictive root is

$$Y^*_{n+1} - \hat{Y}^*_{n+1} = \hat{\sigma}(y_n, \ldots, y_{n-p+1})\varepsilon^*_{n+1} + A^*_\mu \tag{7.11}$$

where $\hat{\sigma}(\cdot)$ is the (consistent) estimator of $\sigma(\cdot)$ that is employed in the bootstrap data generation mechanism. Hence, the following definition is immediate.

Definition 7.2.2 ***Asymptotic pertinence of bootstrap prediction intervals under model*** (7.2). *Consider a bootstrap prediction interval for Y_{n+1} that is based on approximating the distribution of the predictive root $Y_{n+1} - \hat{Y}_{n+1}$ of eq.* (7.10) *by the distribution of the bootstrap predictive root $Y^*_{n+1} - \hat{Y}^*_{n+1}$ of eq.* (7.11). *The interval will be called* asymptotically pertinent *provided the bootstrap satisfies conditions (i)—(iii) of Definition 7.2.1 together with the additional requirement:*

(iv′) $\sigma(y_n, \ldots, y_{n-p+1}) - \hat{\sigma}(y_n, \ldots, y_{n-p+1}) \xrightarrow{P} 0$.

Furthermore, the bootstrap prediction interval for Y_{n+1} that is based on the approximating the distribution of the studentized *predictive root $(Y_{n+1} - \hat{Y}_{n+1})/\hat{V}_n$ by the distribution of the bootstrap studentized predictive root $(Y^*_{n+1} - \hat{Y}^*_{n+1})/\hat{V}^*_n$ will be called asymptotically pertinent if, in addition condition (iv′) of Definition 7.2.1 also holds.*

of (not or)

7.3.1 Forward Bootstrap with Fitted Residuals

Given a sample $Y_1 = y_1, \ldots, Y_n = y_n$ from (7.12), the following are the steps needed to construct the prediction interval for future value Y_{n+1} based on the predictive root method.

Algorithm 7.3.1 FORWARD BOOTSTRAP WITH FITTED RESIDUALS (FF)

1. *Use all observations $y_1, \cdots, y_n$ to obtain the Least Squares (LS) estimators $\hat{\underline{\phi}} = (\hat{\phi}_0, \hat{\phi}_1, \cdots, \hat{\phi}_p)'$ by fitting the following linear model*

$$\begin{pmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{p+1} \end{pmatrix} = \begin{bmatrix} 1 & y_{n-1} & \cdots & y_{n-p} \\ 1 & y_{n-2} & \cdots & y_{n-p-1} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & y_p & \cdots & y_1 \end{bmatrix} \begin{pmatrix} \phi_0 \\ \phi_1 \\ \vdots \\ \phi_p \end{pmatrix} + \begin{pmatrix} \varepsilon_n \\ \varepsilon_{n-1} \\ \vdots \\ \varepsilon_{p+1} \end{pmatrix}. \tag{7.13}$$

2. *For $t = p+1, \cdots, n$, compute the fitted value and fitted residuals:*

$$\hat{y}_t = \hat{\phi}_0 + \sum_{j=1}^{p} \hat{\phi}_j y_{t-j}, \text{ and } \hat{\varepsilon}_t = y_t - \hat{y}_t.$$

3. *Center the fitted residuals: let $r_t = \hat{\varepsilon}_t - \bar{\hat{\varepsilon}}$ for $t = p+1, \cdots, n$, and $\bar{\hat{\varepsilon}} = (n-p)^{-1} \sum_{p+1}^{n} \hat{\varepsilon}_t$; let the empirical distribution of r_t be denoted by $\hat{F}_n$.*

   (a) *Draw bootstrap pseudo-residuals $\varepsilon_1^*, \varepsilon_2^*, \ldots$ i.i.d. from $\hat{F}_n$.*

   (b) *To ensure stationarity of the bootstrap series, we can use an arbitrary initial condition such as $(u_1^*, \cdots, u_p^*) = (0, \ldots, 0)$, generate $n+M$ pseudo-data for some large positive integer M, and then discard the first M data. In other words, generate $\{u_t^*, t \geq p+1\}$ by the recursion:*

$$u_t^* = \hat{\phi}_0 + \sum_{j=1}^{p} \hat{\phi}_j u_{t-j}^* + \varepsilon_t^* \text{ for } t = p+1, \cdots, n+M.$$

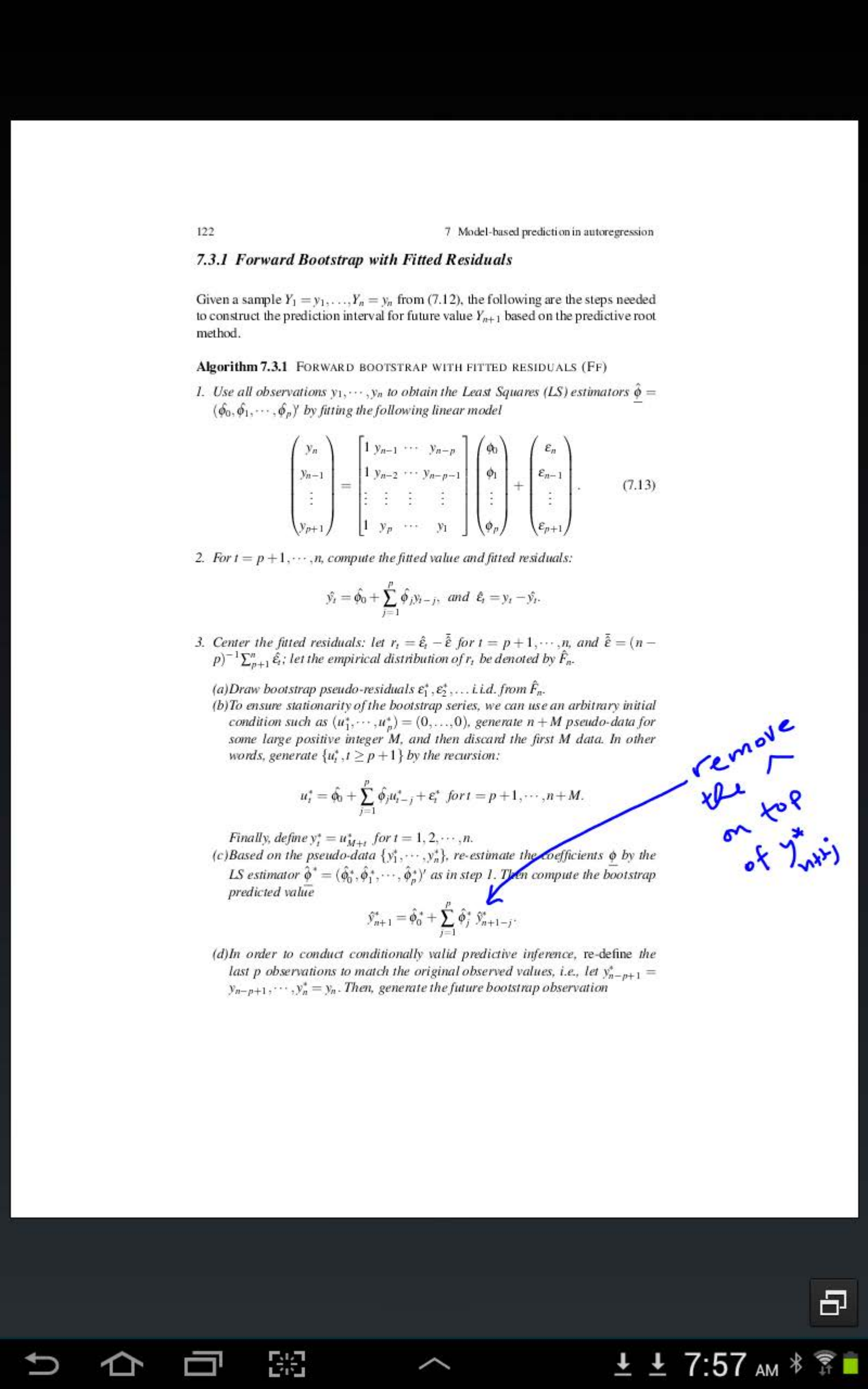
   Finally, define $y_t^* = u_{M+t}^*$ for $t = 1, 2, \cdots, n$.

   (c) *Based on the pseudo-data $\{y_1^*, \cdots, y_n^*\}$, re-estimate the coefficients $\underline{\phi}$ by the LS estimator $\hat{\underline{\phi}}^* = (\hat{\phi}_0^*, \hat{\phi}_1^*, \cdots, \hat{\phi}_p^*)'$ as in step 1. Then compute the bootstrap predicted value*

$$\hat{y}_{n+1}^* = \hat{\phi}_0^* + \sum_{j=1}^{p} \hat{\phi}_j^* \, \hat{y}_{n+1-j}^*.$$

   (d) *In order to conduct conditionally valid predictive inference,* re-define *the last p observations to match the original observed values, i.e., let $y_{n-p+1}^* = y_{n-p+1}, \cdots, y_n^* = y_n$. Then, generate the future bootstrap observation*

remove the ^ on top of y^*_{n+1-j}



Replace two occurrences of h by 1
i.e. $y^*_{n+1} - \hat{y}^*_{n+1}$

remove the ⌜ on top of y_{n+1-j}

$$y^*_{n+1} = \hat{\phi}_0 + \sum_{j=1}^{p} \hat{\phi}_j y^*_{n+1-j} + \varepsilon^*_{n+1}.$$

(e) Calculate a bootstrap root replicate as $y^*_{n+h} - \hat{y}^*_{n+h}$.

4. *Steps (a)-(e) above are repeated B times, and the B bootstrap replicates are collected in the form of an empirical distribution whose* α*-quantile is denoted* $q(\alpha)$.
5. *Compute the predicted value* $\hat{y}_{n+1} = \hat{\phi}_0 + \sum_{j=1}^{p} \hat{\phi}_j \hat{y}_{n+1-j}$.
6. *Construct the* $(1-\alpha)100\%$ *equal-tailed prediction interval for* Y_{n+1} *as*

$$[\hat{y}_{n+1} + q(\alpha/2), \hat{y}_{n+1} + q(1-\alpha/2)]. \tag{7.14}$$

Remark 7.3.1 Step 3 (b) of the above algorithm describes one method to generate a stationary stretch of a time series defined by an autoregressive (or in general Markovian) structure; the technique allows the practitioner to not worry about the initial conditions. A different approach is to generate the starting points of the autoregression from its stationary distribution, e.g., replace Step 3 (b) by (b′) below:

(b′) Let $(y^*_1, \cdots, y^*_p)$ *be chosen at random from the set of p–tuplets* $\{(y_k, \cdots, y_{k+p-1})$ *for* $k = 1, \cdots, n-p+1\}$. *Then, generate* $\{y^*_t, t \geq p+1\}$ *by the recursion:*

$$y^*_t = \hat{\phi}_0 + \sum_{j=1}^{p} \hat{\phi}_j y^*_{t-j} + \varepsilon^*_t \quad for\, t = p+1, \cdots, n.$$

In what follows, we will use either (or both) of these techniques in order to generate stationary autoregressive (or Markovian) time series in the bootstrap world.

Remark 7.3.2 Algorithm 7.3.1 focuses on one-step-ahead prediction for simplicity. However, it is straightforward to extend these results—as well as those in the sequel—in order to construct a prediction interval for Y_{n+h} for some $h \geq 1$ based on the data $\underline{Y}_n$. In addition, the use of resampling affords us the possibility of constructing joint, i.e., simultaneous, prediction intervals for $Y_{n+1}, \ldots, Y_{n+h}$ with prespecified coverage level; details are given in Pan and Politis (2015).

7.3.2 Forward Bootstrap with Predictive Residuals

As in Chapter 3, we may consider using predictive—as opposed to fitted—residuals for the bootstrap. We define the predictive residuals in the AR context as $\hat{\varepsilon}^{(t)}_t = y_t - \hat{y}^{(t)}_t$ where $\hat{y}^{(t)}_t$ is computed from the delete-y_t data set, i.e., the available data for the scatterplot of y_k vs. $\{y_{k-p}, \cdots, y_{k-1}\}$ over which the LS fitting that takes place excludes the single point that corresponds to $k = t$. The forward bootstrap with predictive residuals is similar to Algorithm 7.3.1 except for Step 2.

7.6 Bootstrap Prediction Intervals for Nonparametric Autoregression

In the last several sections, the focus was on prediction intervals for linear autoregressions. In a *nonlinear* autoregression setting, backward bootstrap methods have not been found useful mainly because it is unclear how to generate a nonlinear model such as eq. (7.1) backwards. By contrast, extension of the four forward bootstrap methods to nonlinear—but parametric—autoregressions is straightforward; see Pan and Politis (2015) for details. In what follows, we provide some details on how to employ the forward bootstrap in order to construct bootstrap prediction intervals under a nonparametric autoregression model fitted via kernel smoothing.

7.6.1 Nonparametric Autoregression with i.i.d Errors

In this subsection, we consider a stationary and geometrically ergodic process satisfying eq. (7.1) with the conditional mean function $\mu(\cdot)$ being unknown but assumed smooth. Given a sample $Y_1 = y_1, \cdots, Y_n = y_n$, let $x_t = (y_t, y_{t-1}, \cdots, y_{t-p+1})'$ as before.

Algorithm 7.6.1 FORWARD BOOTSTRAP WITH FITTED RESIDUALS (FF)

1. *For $x \in \mathbf{R}^p$, construct the Nadaraya-Watson kernel estimator $\hat{m}(\cdot)$ as*

$$\hat{m}(x) = \frac{\sum_{t=p}^{n-1} K(\frac{\|x-x_t\|}{h}) y_{t+1}}{\sum_{t=p}^{n-1} K(\frac{\|x-x_t\|}{h})} \tag{7.23}$$

where $\|\cdot\|$ is a norm on $\mathbf{R}^p$, and $K(\cdot)$ is compactly supported, symmetric density function with bounded derivative. As usual, the bandwidth satisfies $h \to 0$ but $hn \to \infty$.

2. *Compute the fitted residuals: $\hat{\varepsilon}_i = y_i - \hat{m}(x_{i-1})$ for $i = p+1, \cdots, n$*
3. *Center the residuals: $\hat{r}_i = \hat{\varepsilon}_i - (n-p)^{-1} \sum_{t=p+1}^{n} \hat{\varepsilon}_t$, for $i = p+1, \cdots, n$.*

   (a) Sample randomly (with replacement) from the values $\hat{r}_{p+1}, \cdots, \hat{r}_n$ to create bootstrap pseudo errors ε_i^* for $i = -M+p, \cdots, n+1$ where M is some large positive number.

   (b) Let $x_p^* = (y_{p+I}, \cdots, y_{1+I})'$ where I is generated as a discrete random variable uniform on the values $0, 1, \ldots, n-p$, and define $(y_{-M}^*, y_{-M+1}^*, \cdots, y_{-M+p-1}^*)' = x_p^*$. Then, generate y_i^* by the recursion:

$$y_i^* = \hat{m}(x_{i-1}^*) + \varepsilon_i^* \quad \text{for } i = -M+p, \cdots, n$$

   where $x_t^* = (y_t^*, \cdots, y_{t-p+1}^*)'$.

   (c) Drop the first M 'burn in' observations to make sure that the starting values have an insignificant effect. Then recompute the kernel estimator $\hat{m}^*(\cdot)$ from

$i = p+1, \cdots, n$

(or replace $-M$ by 1)

L_2– and L_1–optimal point predictors of Y_{n+1}. We will give these in detail as part of the general algorithms for the construction of Model-free predictors and prediction intervals.

Algorithm 9.3.1 MODEL-FREE (MF) POINT PREDICTORS AND PREDICTION INTERVALS FOR Y_{n+1}

1. *Construct $U_1,\ldots,U_n$ by Eq. (9.25) with $D_t(\cdot)$ estimated by either $\bar{D}_t(\cdot)$ or $\bar{D}_t^{LL}(\cdot)$; for the latter, use the respective formulas with $T=t$.*
2. *Construct $Z_1,\ldots,Z_n$ by Eq. (9.26), and use the methods of Sect. 9.3.4 to estimate Γ_n by either $\hat{\Gamma}_n^{AR}$ or $\hat{\Gamma}_n^{\star}$.*
3. *Construct $\varepsilon_1,\ldots,\varepsilon_n$ by Eq. (9.27), and let $\hat{F}_n$ denote their empirical distribution.*
4. *The Model-free L_2–optimal point predictor of Y_{n+1} is then given by*

$$\hat{Y}_{n+1} = \int g_{n+1}(x)d\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} g_{n+1}(\varepsilon_i) \tag{9.38}$$

   where the function g_{n+1} is defined in the predictive equation (9.37) with $D_{n+1}(\cdot)$ being again estimated by either $\bar{D}_{n+1}(\cdot)$ or $\bar{D}_{n+1}^{LL}(\cdot)$, both with $T=t$.
5. *The Model-free L_1–optimal point predictor of Y_{n+1} is given by the median of the set $\{g_{n+1}(\varepsilon_i)$ for $i=1,\ldots,n\}$.*
6. *Prediction intervals for Y_{n+1} with prespecified coverage probability can be constructed via the Model-free Bootstrap of Algorithm 2.4.1 based on either the L_2– or L_1–optimal point predictor.*

Remark 9.3.3 Note that Eq. (9.38) gives an approximation to the *bona fide* L_2–optimal predictor of Y_{n+1} without resorting to the L_2–optimal *linear* predictor (9.7) as in the model-based case.

Algorithm 9.3.1 used the construction of $\bar{D}_t(\cdot)$ or $\bar{D}_t^{LL}(\cdot)$ with $T=t$; using $T=t-1$ instead, leads to the following *predictive* version of the algorithm.

Algorithm 9.3.2 PREDICTIVE MODEL-FREE (PMF) POINT PREDICTORS AND PREDICTION INTERVALS FOR Y_{n+1}
The algorithm is identical to Algorithm 9.3.1 except for using $T=t-1$ instead of $T=t$ in the construction of $\bar{D}_t(\cdot)$ and $\bar{D}_t^{LL}(\cdot)$.

Remark 9.3.4 Under a model-free setup of a locally stationary time series, Paparoditis and Politis (2002b) proposed the Local Block Bootstrap (LBB) in order to generate pseudo-series $Y_1^*,\ldots,Y_n^*$ whose probability structure mimics that of the observed data $Y_1,\ldots,Y_n$. The Local Block Bootstrap has been found useful for the construction of confidence intervals; see Dowla et al. (2003, 2013). However, it is unclear if/how the LBB can be employed for the construction of predictors and prediction intervals for Y_{n+1}.

Recall that when the theoretical transformation H_n is employed, the variables $\varepsilon_1,\ldots,\varepsilon_n$ are i.i.d. $N(0,1)$. Due to the fact that features of H_n are unknown and

must be estimated from the data, the practically available variables $\varepsilon_1, \ldots, \varepsilon_n$ are only approximately i.i.d. $N(0,1)$. However, their empirical distribution of $\hat{F}_n$ converges to $F = \Phi$ as $n \to \infty$. Hence, it is possible to use the limit distribution $F = \Phi$ in instead of $\hat{F}_n$ in both the construction of point predictors and the prediction intervals; this is an application of the Limit Model-Free (LMF) approach. The LMF Algorithm is simpler than Algorithm 9.3.2 as the first three steps of the latter can be omitted. As a matter of fact, the LMF Algorithm is totally based on the inverse transformation $H_{n+1}^{-1} : \underline{\varepsilon}_{n+1} \mapsto \underline{Y}_{n+1}$; the forward transformation $H_n : \underline{Y}_n \mapsto \underline{\varepsilon}_n$ is not needed at all. But for the inverse transformation it is sufficient to estimate $D_t(y)$ by the step functions $\hat{D}_t(y)$ or $\hat{D}_t^{LL}(y)$ with the understanding that their inverse must be a *quantile* inverse; recall that the quantile inverse of a distribution $D(y)$ is defined as $D^{-1}(\beta) = \inf\{y \text{ such that } D(y) \geq \beta\}$.

Algorithm 9.3.3 LIMIT MODEL-FREE (LMF) POINT PREDICTORS AND PREDICTION INTERVALS FOR Y_{n+1}

1. *The LMF L_2–optimal point predictor of Y_{n+1} is*

$$\hat{Y}_{n+1} = \int x g_{n+1}(x) d\Phi(x) \tag{9.39}$$

   where the function g_{n+1} is defined in the predictive equation (9.37) where $D_{n+1}(\cdot)$ is estimated by either $\hat{D}_{n+1}(\cdot)$ or $\hat{D}_{n+1}^{LL}(\cdot)$, both with $T = t - 1$.
2. *In practice, the integral (9.39) can be approximated by Monte Carlo, i.e.,*

$$\int x g_{n+1}(x) d\Phi(x) \simeq \frac{1}{M} \sum_{i=1}^{M} x_i g_{n+1}(x_i) \tag{9.40}$$

   where $x_1, \ldots, x_M$ are generated as i.i.d. $N(0,1)$, and M is some large integer.
3. *Using the above Monte Carlo framework, the LMF L_1–optimal point predictor of Y_{n+1} can be approximated by the median of the set $\{g_{n+1}(x_i)$ for $i = 1, \ldots, M\}$.*
4. *Prediction intervals for Y_{n+1} with prespecified coverage probability can be constructed via the LMF Bootstrap of Algorithm 2.4.3 based on either the L_2– or L_1–optimal point predictor.*

Remark 9.3.5 Interestingly, there is a closed-form (approximate) solution for the LMF L_1–optimal point predictor of Y_{n+1} that can also be used in Step 5 of Algorithm 9.3.1. To elaborate, first note that under the assumed weak dependence, e.g., strong mixing, of the series $\{Y_t\}$ (and therefore also of $\{Z_t\}$), we have the following approximations (for large n), namely:

$$Median\,(Z_{n+1}|\mathscr{F}_1^n(Z)) \simeq Median\,\left(Z_{n+1}|\mathscr{F}_{-\infty}^n(Z)\right)$$

$$= Median\,\left(Z_{n+1}|\mathscr{F}_{-\infty}^n(Y)\right) \simeq Median\,(Z_{n+1}|\mathscr{F}_1^n(Y)).$$

S. Das and D.N. Politis, Predictive inference for locally stationary time series with an application to climate data, *J.Amer.Statist.Assoc.*, vol. 116, no. 534, pp. 919-934, 2021.

[FOR UPDATED/CORRECTED VERSION]