Optimal Linear Interpolation of Multiple Missing Values

Tucker S. McElroy*
Research and Methodology Directorate
U.S. Census Bureau
4600 Silver Hill Road
Washington, D.C. 20233-9100
tucker.s.mcelroy@census.gov

Dimitris N. Politis
Department of Mathematics
and Halicioglu Data Science Institute
University of California, San Diego
La Jolla, CA 92093-0112, USA
dpolitis@ucsd.edu

Abstract

The problem of linear interpolation in the context of a multivariate time series having multiple (possibly non-consecutive) missing values is studied. A concise formula for the optimal interpolating filter is derived, and illustrations using two simple models are provided.

Keywords: Imputation, infinite past, linear filter, prediction.

1 Introduction

In the setting of a covariance-stationary, mean zero time series $\{X_t, t \in \mathbb{Z}\}$, the standard linear prediction problem amounts to extrapolating the one-step ahead (or multiple steps ahead) value based on the observed past; this is typically done via orthogonal projection of the random variable of interest on the linear span of the observed data—see e.g. Ch. 2 of Brockwell and Davis (1991). The interpolation problem is closely related but now we assume that the entire time series has been observed except one particular value. Because of stationarity, and without loss of generality, we may assume that the missing value is X_0, i.e., the data consist of $\{X_t, t \in \mathbb{Z} \setminus \{0\}\}$. Then, the goal is to 'predict' X_0 as a linear function of the data $\{X_t, t \in \mathbb{Z} \setminus \{0\}\}$; this is also done via orthogonal projection—see Wiener (1949) who pioneered the projection technique.

The one-step ahead prediction problem has as its solution the autoregressive coefficients of the process; see e.g. Ch. 6.1 of McElroy and Politis (2020). By contrast, the optimal interpolation problem leads to the notion of *inverse autocorrelations*; see e.g. Politis (1992). Cleveland (1972) apparently coined this term, but the solution to the interpolation problem was developed much

*This report is released to inform interested parties of research and to encourage discussion. The views expressed on statistical issues are those of the authors and not those of the U.S. Census Bureau.

earlier; see Chapter 2.3 of Grenander and Rosenblatt (1957), as well as Wiener's (1949) original monograph. Working concurrently in the Soviet Union, Kolmogorov (1941) also studied the interpolation problem. Later developments include Masani (1960), Chapter 2 of Rozanov (1967), Salehi (1979), Franke (1984), Pourahmadi (1989), and Cheng and Pourahmadi (1997). More recently, further results have been obtained by Bondon (2005), Kasahara et al. (2009), and Inoue (2021); Lepot et al. (2017) provides a recent review.

Interpolation and extrapolation have been also well-studied in the context of a multivariate time series; see Ch. 6 of Hannan (1970), as well as some of the above references, such as Masani (1960). In the context outlined above interpolation results have only been obtained for the case of a single missing value or consecutive stretches of missing values – as discussed in Inoue (2021). (There is an extensive literature treating the missing value problem for finite-length samples, for example using the Kalman filter, but in this note we are focused on infinite-length samples.) The key novelty of our work, with respect to prior literature, is to provide interpolation formulas for the case of general, possibly non-consecutive, patterns of missing values present in multivariate time series. In Section 2 we derive the form of the optimal linear interpolator in the general multivariate case; application to the univariate case is immediate. Section 3 shows the validity of the optimal interpolating filter in the case of possibly slow decay of the autocovariances. Illustrations using two simple models are provided in Section 4. Appendix A presents some background on optimal linear prediction in the multivariate case, while Appendix B contains the technical proofs.

Remark 1 The objective of our paper is to derive the formula for the mean squared error optimal linear interpolator with multiple missing values; this is a theoretical result without regard to any data, following the Wiener-Kolmogorov theory of linear prediction as Hilbert space projection. Having derived the formula of the optimal filter, its coefficients could indeed be estimated from data by one of the usual methods, e.g., Method of Moments, Least Squares, pseudo Maximum Likelihood, etc.

2 Optimal Linear Interpolation of Multiple Missing Values

Consider a covariance-stationary, mean zero, *multivariate* time series $\{X_t, t \in \mathbb{Z}\}$ with autocovariance at lag h denoted by $\gamma_h = \mathbb{E}X_t X_{t-h}'$ where $'$ denotes the transpose; each X_t is an N-dimensional column vector, and therefore γ_h is an $N \times N$ matrix. A typical assumption is that the *autocovariance generating function* $\gamma(z) = \sum_{h\in\mathbb{Z}} \gamma_h z^h$ is well-defined for z in some annulus about the unit circle of the complex plane; this converging power series in both powers of z and z^{-1} is called a Laurent series, and characterizes a *holomorphic* function of a complex variable. In this case, the *spectral density* of the time series $\{X_t\}$ is given by the function $f(\lambda) = \gamma(e^{i\lambda})$ for $\lambda \in [-\pi, \pi]$.

If the determinant of $\gamma(z)$ has no zeroes on the unit circle, then the function $\xi(z) = \gamma(z)^{-1}$ is also holomorphic on some annulus about the unit circle. Consequently, it has a converging

Laurent series expansion $\xi(z) = \sum_{h\in\mathbb{Z}} \xi_h z^h$ for some coefficients $\{\xi_h\}_{h\in\mathbb{Z}}$ that we will call the *inverse autocovariances*; therefore, $\xi(z)$ is the inverse autocovariance generating function. We can also define the *inverse autocorrelation* at lag h by $\zeta_h = \xi_0^{-1}\xi_h$; see also Ch. 6.3 of McElroy and Politis (2020).

Our goal is to derive a linear filter to predict a finite number of missing values that are assumed to be completely missing (i.e., either all or none of the components of each random vector are available)[1]. There will be one filter for each missing value, which is viewed as the target – just like there is one filter for each forecast lead in a multi-step ahead forecasting problem.

We need a reference time t, in terms of which to formulate the filters. Because of stationarity, this is not too important; hence, for ease of exposition, we choose $t = 0$ as the reference point. Then, the missing value times will be denoted as $k \in M$; we assume to have m missing values where $m = |M| < \infty$ (here $|M|$ denotes the number of elements in the set M). The elements of M are indexed in some way, and can be written $M = \{j_1, j_2, \ldots, j_m\}$. Without loss of generality, suppose these are ordered such that $j_1 < j_2 < \ldots < j_m$ (these integers are of course distinct).

Fix $k \in M$. The dataset available to predict the specific missing value X_k is $\{X_j\}_{j\notin M}$. To predict X_k from $\{X_j\}_{j\notin M}$ we can employ the linear predictor

$$\widehat{X}_k = \sum_{j\notin M} \pi_j X_j. \tag{1}$$

The filter coefficients π_j are $N \times N$-dimensional matrices; they depend on k, but we suppress this to facilitate the notation. The challenge is to identify the *optimal* linear predictor, i.e., identify the optimal coefficients π_j that minimize the Mean Square Error (MSE) of prediction[2] that is defined as $\mathbb{E}(X_{t-k} - \widehat{X}_{t-k})(X_{t-k} - \widehat{X}_{t-k})'$.

Note that the filter (1) does not rely on any values that are missing. Hence, we can re-write the predictor (1) as

$$\widehat{X}_k = \sum_{j\in\mathbb{Z}} \pi_j X_j \tag{2}$$

with the understanding that $\pi_j = 0$ for $j \in M$; we call these the *zero constraints* of the filter. Apparently, (2) can be expressed as $\widehat{X}_k = \pi(B^{-1})X_0$, where B is the backshift operator (and B^{-1} is a forward shift), i.e., $BX_t = X_{t-1}$. The function $\pi(z)$ is then defined via

$$\pi(z) = \sum_{j\in\mathbb{Z}} \pi_j z^j = \sum_{j\notin M} \pi_j z^j. \tag{3}$$

In our main result below, we identify the function $\pi(z)$ associated with the optimal linear filter. Note that knowledge of $\pi(z)$ for all z in an annulus about the unit circle is tantamount to knowing

[1]This assumption precludes the handling of mixed frequency data – where a lower frequency section of the time series is viewed as a higher frequency time series observed with missing values – as well as ragged edge data (where the vector components at a particular time point may be partially missing).

[2]Minimization of the MSE matrix is considered in the usual way, in the sense of an ordering of non-negative definite matrices; see Appendix A.

the coefficient sequence $\{\pi_j, j \in \mathbb{Z}\}$; to see why, we can plug $z = e^{-i\lambda}$ (where $i = \sqrt{-1}$) into (3), yielding the Fourier series

$$\pi(e^{-i\lambda}) = \sum_{j\in\mathbb{Z}} \pi_j e^{-ij\lambda}. \tag{4}$$

Consequently, the j–th Fourier coefficient of the function $\pi(e^{-i\lambda})$ is given by

$$\pi_j = (2\pi)^{-1} \int_{-\pi}^{\pi} \pi(e^{-i\lambda}) e^{ij\lambda} d\lambda. \tag{5}$$

To facilitate notation, in what follows we will denote $\langle g(z) \rangle = (2\pi)^{-1} \int_{-\pi}^{\pi} g(e^{-i\lambda}) d\lambda$; we will also employ the following assumption:

Assumption A. Assume that the autocovariance generating function $\gamma(z) = \sum_{h\in\mathbb{Z}} \gamma_h z^h$ is well-defined for z in some annulus about the unit circle of the complex plane. Further assume that the determinant of $\gamma(z)$ has no zeroes on the unit circle, i.e., that $\gamma(z) \neq 0$ whenever $|z| = 1$.

Theorem 1 *Assume Assumption A. Then, we can compute the inverse autocovariances via*

$$\xi_h = \langle z^{-h} \gamma(z)^{-1} \rangle. \tag{6}$$

Furthermore, the MSE–optimal linear predictor of X_k given data $\{X_j\}_{j\notin M}$, where $M = \{j_1, j_2, \ldots, j_m\}$, is given by (1). Letting $k = j_u$, and setting $\pi_{j_r} = 0$ for $1 \leq r \leq m$, the formula for $\pi(z) = \sum_{j\in\mathbb{Z}} \pi_j z^j$ is given by

$$\pi(z) = I z^k - E_u \, \Xi^{-1} \, Z \, \xi(z^{-1}), \tag{7}$$

where $E_u = \underline{e}_u' \otimes I$. Here, I denotes the $N \times N$ identity matrix, $\otimes$ denotes Kronecker product, $\underline{e}_u$ is a dimension m unit vector (with one in position u for $1 \leq u \leq m$, and zero else), Ξ is a $mN \times mN$-dimensional block Toeplitz matrix with rs–th block entry (for $1 \leq r, s \leq m$) given by $\xi_{j_r - j_s}$, and $Z = \underline{z} \otimes I$, where $\underline{z}$ is a vector with entries z^{j_r} for $1 \leq r \leq m$.

Finally, the optimal predictor's MSE matrix equals $E_u \, \Xi^{-\prime} \, E_u'$, where $-\prime$ denotes inverse transpose.

Remark 2 The filter formula (7) automatically generates the zero constraints, because the jth coefficient is given by

$$\pi_j = I 1_{\{j=k\}} - \sum_{r=1}^{m} \Xi_{u,r}^{-1} \, \xi_{j_r - j}, \tag{8}$$

where $1_{\{j=k\}}$ is the indicator of set $\{j = k\}$. If $j \in M$ then $j = j_v$ for some $1 \leq v \leq m$, implying

$$\pi_j = 1_{\{v=u\}} - \sum_{r=1}^{m} \Xi_{u,r}^{-1} \, \Xi_{r,v} = 1_{\{v=u\}} - \left(\Xi^{-1}\Xi\right)_{u,v} = 0$$

(since $\xi_{j_r - j_v} = \Xi_{r,v}$). The pattern of the non-zero weights depends on whether the missing values are clustered (i.e., contiguous) or isolated from one another, as this impacts the structure of Ξ.

The results are somewhat easier to state in the univariate case ($N = 1$), as summarized in the following corollary. Note that in this case the autocovariance generating function $\gamma(z)$, as well as its inverse $\gamma(z)^{-1}$, are real-valued (scalar) functions. Historically, this univariate case has been of great interest; Masani (1960), Rozanov (1967), and Inoue (2021) discuss the situation where all the missing values are consecutive.

Corollary 1 *Assume Assumption A, and that $N = 1$, i.e., the univariate case. Then, we can compute the inverse autocovariances via (6) as before.*

The MSE–optimal linear predictor of X_k given data $\{X_j\}_{j \notin M}$, where $M = \{j_1, j_2, \ldots, j_m\}$, is given by (1). Letting $k = j_u$, and setting $\pi_{j_r} = 0$ for $1 \leq r \leq m$, the formula for $\pi(z) = \sum_{j \in \mathbb{Z}} \pi_j z^j$ is given by

$$\pi(z) = z^k - \underline{e}_u' \, \Xi^{-1} \, \underline{z} \, \xi(z^{-1}), \tag{9}$$

where $\underline{e}_u$ is a dimension m unit vector (with one in position u for $1 \leq u \leq m$, and zero else), Ξ is a $m \times m$-dimensional Toeplitz matrix with rs–th entry (for $1 \leq r, s \leq m$) given by $\xi_{j_r - j_s}$, and $\underline{z}$ is a vector with entries z^{j_r} for $1 \leq r \leq m$.

Finally, the optimal predictor's MSE equals $\underline{e}_u' \, \Xi^{-\prime} \, \underline{e}_u$, where $-\prime$ denotes inverse transpose.

3 Optimal Linear Interpolation in the Case of a Possible Slow Decay of the Autocovariances

In the previous section, it was assumed that the autocovariance generating function $\gamma(z)$ is well-defined over an annulus about the unit circle. However, this presupposes that the autocovariances γ_h decay to zero exponentially fast as the lag h increases. While this is true in many cases, e.g. in stationary ARMA models, there are certainly situations where γ_h only decays slowly in h, with a polynomial rate. Even in such a case, the optimal interpolating filter would still be given by eq. (7) of Theorem 1 *provided the infinite sums implicit in this formula are well-defined.*

Recall that the coefficients of the optimal filter can be computed via the values of $\gamma(z)$ for z on the unit circle; see eq. (4) and (5). Hence, we may focus our attention to the spectral density $f(\lambda) = \gamma(e^{-i\lambda})$, and try to find a sufficient condition for the optimal filter (7) to be well-defined. To this end, we are aided by *Wiener's lemma*; see Wiener (1932) for the original, and Sun (2007) or Shin and Sun (2013) for extensions and elaborations.

In the univariate case ($N = 1$), the classical Wiener's lemma states that if a periodic function $f(\lambda)$ has an absolutely convergent Fourier series and never vanishes, then $1/f(\lambda)$ also has an absolutely convergent Fourier series. We will formulate a multivariate version of Wiener's lemma that will help us define the inverse autocovariances when $\gamma(z)$ is well-defined only on the unit circle. Let $\| \cdot \|_2$ denote the matrix 2-norm[3]; we will require the following assumption:

[3] Alternatively, we can use the matrix Frobenius norm in the result's formulation.

Assumption B. Assume that the autocovariances γ_h have summable matrix 2-norm, i.e., $\sum_{h\in\mathbb{Z}} \|\gamma_h\|_2 < \infty$, and that the determinant of $f(\lambda)$, denoted by $\det f(\lambda)$, does not vanish for $\lambda \in [-\pi, \pi]$.

The following is a multivariate version of Wiener's lemma that is sufficient for our purposes. The result is not necessarily new in view of the general Banach space theorems of Bochner and Phillips (1942). Nevertheless, for completeness and concreteness, we provide a statement and proof in the finite-dimensional matrix setting.

Lemma 1 *Suppose that $f(\lambda) = \gamma(e^{-i\lambda})$ satisfies Assumption B. Then the inverse $f(\lambda)^{-1}$ is well-defined for $\lambda \in [-\pi, \pi]$, and has Fourier coefficients $\{\xi_h\}$ with summable matrix 2-norm.*

Under the conditions of the above Lemma, the function $\xi(e^{-i\lambda}) = \gamma(e^{-i\lambda})^{-1}$ will have an absolutely convergent Fourier series $\xi(e^{-i\lambda}) = \sum_{h\in\mathbb{Z}} \xi_h e^{-ih\lambda}$. The coefficients $\{\xi_h\}_{h\in\mathbb{Z}}$ are the *inverse autocovariances*; note that this is identical to eq. (6). The inverse autocorrelations are defined by $\zeta_h = \xi_0^{-1}\xi_h$ as before. In view of Lemma 1, the following is immediate.

Theorem 2 *Assume Assumption B instead of Assumption A. Then, the conclusions and formulas of Theorem 1 remain valid.*

Remark 3 It is also shown in the proof of Lemma 1 that the absolute value of each component of both f and f^{-1} has an absolutely convergent Fourier series – and in particular is absolutely integrable over $[-\pi, \pi]$ – hence impying that (1.2) and (1.3) of Inoue (2021) hold. Therefore Assumption B implies "minimality", a concept defined by Kolmogorov – see Masani (1960), which essentially states that X_t (for any $t \in \mathbb{Z}$) cannot be perfectly predicted from all the other random variables $\{X_s\}_{s\neq t}$ of the stochastic process. Hence, Theorem 2 could also be proved by assuming minimality instead of Assumption B. However, we note that Assumption B is a weak assumption that is convenient to verify. For example, in the univariate case, assuming absolute summability of the autocovariances is easily checked, and moreover is necessary to ensure that the spectral density f is well-defined.

For completeness, we also state the corollary in the univariate case.

Corollary 2 *Assume $N = 1$. Instead of Assumption A, assume that the autocovariances γ_h are absolutely summable, and that $f(\lambda)$ does not vanish for $\lambda \in [-\pi, \pi]$. Then, the conclusions and formulas of Corollary 1 remain valid.*

4 Illustrations

In this section we consider two illustrations: a VAR(1) and a VMA(1), i.e., a Vector Autoregressive model of order 1, and a Vector Moving Average model of order 1; these are both examples where

the autocovariance generating function is well-defined over an annulus. Within the multivariate examples, we also examine the case of $N = 1$ as well.

4.1 VAR(1)

Suppose that $\{X_t\}$ is a VAR(1) process with coefficient matrix Φ and innovation covariance Σ, i.e., $X_t = \Phi X_{t-1} + Z_t$ where Z_t is a mean zero white noise with $\Sigma = \mathbb{E} Z_t Z_t'$. In this case,

$$\begin{aligned}
\gamma(z) &= (I - \Phi z)^{-1} \Sigma (I - \Phi' z^{-1})^{-1} \\
\xi(z) &= (I - \Phi' z^{-1}) \Sigma^{-1} (I - \Phi z) \\
\xi_0 &= \Sigma^{-1} + \Phi' \Sigma^{-1} \Phi \\
\xi_1 &= -\Sigma^{-1} \Phi.
\end{aligned}$$

All but three of the inverse autocovariances are zero, and hence the summation in (8) can involve at most three terms, involving ξ_{-1}, ξ_0, ξ_1. The structure of Ξ is

$$\Xi = \begin{bmatrix}
\xi_0 & \xi_{j_1 - j_2} & 0 & \dots & 0 \\
\xi_{j_2 - j_1} & \xi_0 & \xi_{j_2 - j_3} & 0 & \dots \\
0 & \xi_{j_3 - j_2} & \xi_0 & \xi_{j_3 - j_4} & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \dots & 0 & \xi_{j_m - j_{m-1}} & \xi_0
\end{bmatrix}.$$

Only the block diagonal, the super-diagonal, and the sub-diagonal are non-zero, and some of these entries can even be zero – the matrix is not Toeplitz in general. For instance, on the super-diagonal each entry is an inverse autocovariance based at a lag that is a consecutive difference of elements of M, and is zero unless this lag equals 1. Hence, entries of the super-diagonal are zero unless they correspond to consecutive (clustered) missing values.

To make the illustration more concrete, suppose that $M = \{0, 1, 3\}$, which means there are two consecutive missing values at times 0 and 1, but an isolated missing value at time 3. Then we find

$$\begin{aligned}
\Xi &= \begin{bmatrix}
\xi_0 & \xi_{-1} & 0 \\
\xi_1 & \xi_0 & 0 \\
0 & 0 & \xi_0
\end{bmatrix} \\
\Xi^{-1} &= \begin{bmatrix}
\xi_0^{-1} + \xi_0^{-1} \xi_{-1} S^{-1} \xi_1 \xi_0^{-1} & -\xi_0^{-1} \xi_{-1} S^{-1} & 0 \\
-S^{-1} \xi_1 \xi_0^{-1} & S^{-1} & 0 \\
0 & 0 & \xi_0^{-1}
\end{bmatrix} \\
S &= \xi_0 - \xi_1 \xi_0^{-1} \xi_{-1}.
\end{aligned}$$

Because of this structure of missing values, $\pi_0 = \pi_1 = \pi_3 = 0$; also from (8) and the fact that $\xi_h = 0$ for $|h| > 1$, we find $\pi_j = 0$ if $j > 4$ or $j < -1$. So only π_{-1}, π_2, and π_4 are non-zero, and their

values depends on which of the three missing value filters we are considering. Setting $\zeta_h = \xi_0^{-1}\xi_h$, for $k = 0$ the coefficients are

$$\begin{aligned}\pi_{-1} &= -\left(\xi_0^{-1} + \xi_0^{-1}\xi_{-1}S^{-1}\xi_1\xi_0^{-1}\right)\xi_1 = -(I - \zeta_{-1}\zeta_1)^{-1}\zeta_1 \\ \pi_2 &= \xi_0^{-1}\xi_{-1}S^{-1}\xi_{-1} = \zeta_{-1}\left(I - \zeta_1\zeta_{-1}\right)^{-1}\zeta_{-1} \\ \pi_4 &= 0,\end{aligned}$$

which shows that only observations at either side of the cluster of missing values get weighted; also their weights are different, which makes sense because observation X_1 has more to say about X_0 than does observation X_2, because it is closer (and hence has a higher degree of association) in lag distance.

Special case: $N = 1$**.** In the univariate case, the coefficients simplify to

$$\pi_{-1} = -\frac{\zeta_1}{1-\zeta_1^2} \qquad \pi_2 = \frac{\zeta_1^2}{1-\zeta_1^2},$$

using $\zeta_{-1} = \zeta_1$. Clearly $\pi_2 = -\zeta_1\pi_{-1}$ and hence $|\pi_2| < |\pi_{-1}|$. Next, the coefficients for the $k = 1$ filter are

$$\begin{aligned}\pi_{-1} &= S^{-1}\xi_1\xi_0^{-1}\xi_1 = \zeta_1\left(I - \zeta_{-1}\zeta_1\right)^{-1}\zeta_1 \\ \pi_2 &= -S^{-1}\xi_{-1} = -(I - \zeta_1\zeta_{-1})^{-1}\zeta_{-1} \\ \pi_4 &= 0.\end{aligned}$$

Again there is no contribution from X_4, but we see the weights for π_{-1} and π_2 are similar to those of π_2 and π_{-1} (respectively) in the $k = 0$ filter, only with ζ_1 and ζ_{-1} interchanged. This is appropriate, and in the univariate case the weights exactly correspond. Finally, the coefficients for the $k = 3$ filter are

$$\begin{aligned}\pi_{-1} &= 0 \\ \pi_2 &= -\xi_0^{-1}\xi_1 = -\zeta_1 \\ \pi_4 &= -\xi_0^{-1}\xi_{-1} = -\zeta_{-1},\end{aligned}$$

which provides equal weights in the univariate case for the observations on either side of the isolated missing value X_3. As a final remark, it is simple to see that the weight π_j does not depend on Σ when $N = 1$, because it cancels out, but when $N > 1$ the matrix Σ cannot be manipulated in this way, and the filter coefficients will depend on Σ.

4.2 VMA (1)

Suppose that $\{X_t\}$ is a VMA(1) process written with a minus convention (for convenience in the following formulas): $X_t = Z_t - \Theta' Z_{t-1}$, where $\{Z_t\}$ is a mean zero white noise with covariance

matrix Σ. This is an atypical parameterization of a VMA(1), where we have written $-\Theta'$ where one would usually expect to see Θ, but this is only done so that the following formulas are less cluttered. It follows that

$$\begin{aligned}\gamma(z) &= (I - \Theta' z)\,\Sigma\,(I - \Theta z^{-1}) \\ \xi(z) &= (I - \Theta z^{-1})^{-1}\,\Sigma^{-1}\,(I - \Theta' z)^{-1} \\ \xi_h &= \sum_{k\geq 0} \Theta^{h+k}\Sigma^{-1}\Theta'^{k} = \Theta^h \xi_0 \quad h \geq 0.\end{aligned}$$

As a result the structure of Ξ is

$$\Xi = \begin{bmatrix} \xi_0 & \xi_0\Theta^{j_2-j_1\prime} & \xi_0\Theta^{j_3-j_1\prime} & \dots \\ \Theta^{j_2-j_1}\xi_0 & \xi_0 & \xi_0\Theta^{j_3-j_2\prime} & \dots \\ \vdots & \ddots & \ddots & \ddots \\ \dots & \Theta^{j_m-j_{m-2}}\xi_0 & \Theta^{j_m-j_{m-1}}\xi_0 & \xi_0 \end{bmatrix}.$$

This block matrix is highly structured, but in general is not block Toeplitz.

For a concrete illustration, suppose that $M = \{0, \ell\}$ for some $\ell > 0$, which means there are two missing values that are clustered when $\ell = 1$. Then we find

$$\begin{aligned}\Xi &= \begin{bmatrix} \xi_0 & \xi_0\Theta^{\ell\prime} \\ \Theta^{\ell}\xi_0 & \xi_0 \end{bmatrix} \\ \Xi^{-1} &= \begin{bmatrix} (I + \zeta_{-\ell}[I - \zeta_\ell\zeta_{-\ell}]^{-1}\zeta_\ell)\xi_0^{-1} & -\zeta_{-\ell}[I - \zeta_\ell\zeta_{-\ell}]^{-1}\xi_0^{-1} \\ -[I - \zeta_\ell\zeta_{-\ell}]^{-1}\zeta_\ell\xi_0^{-1} & [I - \zeta_\ell\zeta_{-\ell}]^{-1}\xi_0^{-1} \end{bmatrix} \\ \pi_j &= \begin{cases} I1_{\{j=0\}} - [I - \zeta_{-\ell}\zeta_\ell]^{-1}(\zeta_{-j} - \zeta_{-\ell}\zeta_{\ell-j}) & \text{if } k = 0 \\ I1_{\{j=\ell\}} - [I - \zeta_\ell\zeta_{-\ell}]^{-1}(\zeta_{\ell-j} - \zeta_\ell\zeta_{-j}) & \text{if } k = \ell. \end{cases}\end{aligned}$$

We can directly see that these coefficients satisfy the zero constraints.

Special case: $N = 1$**.** In the univariate case, we can make some further simplifications, because $\zeta_h = \theta^{|h|}$ writing the scalar θ for Θ'; recall that in the multivariate case $\zeta_h \neq \zeta_{-h}$ in general. If $k = 0$, then

$$\pi_j = \begin{cases} 1_{\{j=0\}} - \theta^{-j} & \text{if } j \leq 0 \\ -(1 - \theta^{2\ell})^{-1}(\theta^j - \theta^{2\ell-j}) & \text{if } 0 < j < \ell \\ 0 \qquad \text{if } j \geq \ell, \end{cases}$$

and if $k = \ell$ then

$$\pi_j = \begin{cases} 1_{\{j=\ell\}} - \theta^{j-\ell} & \text{if } j \geq \ell \\ -(1 - \theta^{2\ell})^{-1}(\theta^{\ell-j} - \theta^{\ell+j}) & \text{if } 0 < j < \ell \\ 0 \qquad \text{if } j \leq 0. \end{cases}$$

One interesting feature is that the coefficients truncate on one side of the filter, past where the second missing value is placed – this differs from the case of a single missing value, where all coefficients (except π_0) are nonzero.

Acknowledgment. Many thanks are due to Qiyu Sun for his helpful suggestions with regards to the multivariate Wiener's lemma. We are also grateful to two anonymous referees for helpful suggestions. Research of the second author was partially supported by NSF grant DMS 19-14556.

Conflict of Interest. On behalf of all authors, the corresponding author states that there is no conflict of interest.

Data Availability. There is no data or code for this article.

References

[1] Artin, M. (1991). *Algebra*, Prentice-Hall, New Jersey.

[2] Bochner, S. and Phillips, R.S. (1942). Absolutely convergent Fourier expansions for non-commutative normed rings, *Ann. Math.*, vol. 43, pp. 409-418.

[3] Bondon, P. (2005). Influence of missing values on the prediction of a stationary time series. *Journal of Time Series Analysis*, vol. 26, no. 4, pp.519-525.

[4] Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods, 2nd ed.*, Springer, New York.

[5] Cheng, R. and Pourahmadi, M. (1997). Prediction with incomplete past and interpolation of missing values, *Statistics & Probability Letters*, vol. 33, no. 4, pp.341-346.

[6] Cleveland, W.S. (1972). The inverse autocorrelations of a time series and their applications, *Technometrics*, vol. 14, no. 2, pp. 277-293.

[7] Franke, J. (1984). On the robust prediction and interpolation of time series in the presence of correlated noise. *Journal of Time Series Analysis*, vol. 5, no. 4, pp.227-244.

[8] Grenander, U. and Rosenblatt, M. (1957). *Statistical Analysis of Stationary Time Series*, Wiley, New York.

[9] Hannan, E.J. (1970). *Multiple Time Series*, John Wiley, New York.

[10] Inoue, A. (2021). Explicit formulas for the inverses of Toeplitz matrices, with applications. arXiv preprint arXiv:2105.01165.

[11] Kasahara, Y., Pourahmadi, M. and Inoue, A. (2009). Duals of random vectors and processes with applications to prediction problems with missing values, *Statistics & Probability Letters*, vol. 79, no. 14, pp.1637-1646.

[12] Kolmogorov, A.N. (1941). Stationary sequences in Hilbert space, *Byul. Moskov. Gos. Univ. Ser. Mat*, vol. 2, no. 6, pp.3-40.

[13] Lepot, M., Aubin, J.B. and Clemens, F.H. (2017). Interpolation in time series: An introductive overview of existing methods, their performance criteria and uncertainty assessment, *Water*, vol. 9, no. 10, 796.

[14] Masani, P. (1960). The prediction theory of multivariate stochastic processes, III, *Acta Mathematica*, vol. 104, no. 1-2, pp.141-162.

[15] McElroy, T.S. and Politis, D.N. (2020). *Time Series: A First Course with Bootstrap Starter*, Chapman and Hall/CRC Press, Boca Raton.

[16] Politis, D.N. (1992). Moving Average processes and Maximum Entropy, *IEEE Trans. Info. Theory*, vol. 38, no. 3, pp. 1174-1177.

[17] Pourahmadi, M. (1989). Estimation and interpolation of missing values of a stationary time series. *Journal of Time Series Analysis*, vol. 10, no. 2, pp.149-169.

[18] Rozanov, I.A. (1967). *Stationary random processes,* Holden-Day.

[19] Salehi, H. (1979). Algorithms for linear interpolator and interpolation error for minimal stationary stochastic processes. *The Annals of Probability*, pp.840-846.

[20] Shin, C.E. and Sun, Q. (2013). Wiener's lemma: localization and various approaches, *Appl. Math. J. Chinese Univ.*, vol. 28, no. 4, pp. 465-484.

[21] Sun, Q. (2007). Wiener's lemma for infinite matrices, *Trans. Amer. Math. Soc.*, vol. 359, pp. 3099-3123.

[22] Wiener, N. (1932). Tauberian theorems, *Ann. Math.*, vol. 33, no. 1, pp. 1-100.

[23] Wiener, N. (1949). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: with Engineering Applications,* MIT Press, Cambridge.

Appendix A Minimal MSE Matrix and Optimal Linear Prediction

For symmetric non-negative definite matrices A and B, we say that $A \geq B$ if and only if $A - B$ is non-negative definite, which we can denote as $A - B \geq 0$. A predictor is said to have minimal MSE

matrix Σ if and only if any other predictor has MSE matrix $\widetilde{\Sigma} \geq \Sigma$. Any optimal linear prediction satisfies the normal equations, and hence has minimal MSE matrix, which can be shown as follows.

Let $\widehat{X}$ be the optimal linear predictor of some random vector X. Then,

$$\Sigma = \mathbb{E}(X - \widehat{X})(X - \widehat{X})'$$

by definition of MSE matrix. By the normal equations, the error $X - \widehat{X}$ is orthogonal to all linear functions of the information set (i.e., the random variables used to form the prediction). Hence any other linear predictor $\widetilde{X}$ has MSE matrix

$$\begin{aligned}\widetilde{\Sigma} &= \mathbb{E}(X - \widetilde{X})(X - \widetilde{X})' \\ &= \mathbb{E}\Big((X - \widehat{X}) + (\widehat{X} - \widetilde{X})\Big)\Big((X - \widehat{X}) + (\widehat{X} - \widetilde{X})\Big)' \\ &= \Sigma + \mathbb{E}(\widehat{X} - \widetilde{X})(\widehat{X} - \widetilde{X})',\end{aligned}$$

using the fact that $\widehat{X} - \widetilde{X}$ is a linear function of the information set, and hence is orthogonal to $X - \widehat{X}$. Therefore

$$\widetilde{\Sigma} - \Sigma = \mathbb{E}(\widehat{X} - \widetilde{X})(\widehat{X} - \widetilde{X})',$$

a non-negative definite matrix; hence, $\widetilde{\Sigma} \geq \Sigma$. □

Appendix B Technical Proofs

Proof of Theorem 1. First note that Assumption A implies that $\xi(z) = \gamma(z)^{-1}$ is also holomorphic on some annulus about the unit circle, i.e., it has the converging Laurent series expansion $\xi(z) = \sum_{h\in\mathbb{Z}} \xi_h z^h$; hence, eq. (6) follows. Now, the normal equations for the optimal predictor are

$$\mathbb{E}\widehat{X}_k X_\ell' = \mathbb{E}X_k X_\ell'$$

for all $\ell \notin M$. This is equivalent to

$$\sum_{j\in\mathbb{Z}} \pi_j \, \gamma_{j-\ell} = \gamma_{k-\ell}$$

or

$$\langle z^\ell \pi(z^{-1})\gamma(z)\rangle = \langle z^{\ell-k}\gamma(z)\rangle.$$

Mapping $z \mapsto z^{-1}$ by change of variable and consolidating, we have

$$0 = \langle z^{-\ell}(Iz^k - \pi(z))\gamma(z^{-1})\rangle$$

for all $\ell \notin M$. We know the function $z^{-\ell}(Iz^k - \pi(z))\gamma(z^{-1})$ is a Laurent series, and the above equation says that the ℓth coefficient of $(Iz^k - \pi(z))\gamma(z^{-1})$ is zero unless $\ell \in M$. Hence there exist real $N \times N$-dimensional coefficient matrices $\{c_h\}_{h\in M}$ such that

$$(Iz^k - \pi(z))\gamma(z^{-1}) = \sum_{h\in M} c_h z^h.$$

Solving for $\pi(z)$, we obtain

$$\pi(z) = Iz^k - \sum_{h\in M} c_h z^h\, \xi(z^{-1}). \tag{A.1}$$

Now we apply the zero constraints by integrating (A.1) against $z^{-\ell}$ for each $\ell \in M$:

$$0 = \pi_\ell = \langle z^{-\ell}\pi(z)\rangle = I1_{\{\ell=k\}} - \sum_{h\in M} c_h \langle z^{h-\ell}\xi(z^{-1})\rangle = I1_{\{\ell=k\}} - \sum_{h\in M} c_h\, \xi_{h-\ell}.$$

These are m linear equations in m (matrix) unknowns, and are easily solved. Let $C = [c_{j_1}, c_{j_2}, \ldots, c_{j_m}]$; recalling that $k = j_u$, our system is written

$$C\,\Xi = [0, \ldots, 0, I, 0, \ldots, 0]\,,$$

where the identity matrix is in position u among m slots. Therefore $C = E_u\,\Xi^{-1}$, with invertibility guaranteed because the corresponding spectral density $\xi(e^{-i\lambda})$ is invertible for all λ. Plugging into (A.1) yields (7). Note that $E_u'\Xi^{-1}$ is just the u-th block row of Ξ^{-1}.

To compute the MSE, recall that B is the backshift operator. The prediction error is

$$X_k - \widehat{X}_k = (B^{-k} - \pi(B^{-1}))X_0 = \sum_{h\in M} c_h B^{-h}\, \xi(B^{-1})X_t,$$

which has mean zero with variance matrix

$$\begin{aligned}
&\langle \sum_{h\in M} c_h z^{-h}\, \xi(z) f(z) \xi(z^{-1})' \sum_{\ell\in M} c_\ell' z^\ell \rangle \\
&= \sum_{h,\ell\in M} c_h \langle z^{\ell-h}\xi(z^{-1})'\rangle c_\ell' \\
&= \sum_{r,s=1}^{m} c_{j_r}\xi_{j_r-j_s}c_{j_s}' \\
&= C\,\Xi\,C' = E_u\,\Xi^{-\prime}\,E_u';
\end{aligned}$$

this is the uu–th block matrix of Ξ^{-1}. $\square$

Proof of Lemma 1. The crux of the argument is that for each λ the matrix $f(\lambda)$ is invertible if and only if its determinant is non-zero; but $\det f(\lambda)$ is a scalar quantity, to which we can apply the original Wiener's lemma. Recall the formula

$$f(\lambda)^{-1} = \frac{1}{\det f(\lambda)}\, f(\lambda)^\sharp, \tag{A.2}$$

where $f(\lambda)^\sharp$ is called the adjoint matrix. Note that $f(\lambda)^\sharp$ is well-defined for any λ, because its matrix entries are given by finite sums and products of the entries of $f(\lambda)$; see e.g. Artin (1991). First, we establish that the Fourier coefficients of the adjoint have square summable matrix 2-norm.

Let the kth Fourier coefficient of the adjoint be denoted $\tau_k = \langle z^{-k}\gamma(z)^\sharp\rangle$. Let $\|\cdot\|_F$ denote the Frobenius norm; then, by the Plancherel identity we have

$$\sum_k \|\tau_k\|_F^2 = \langle \|\gamma(z)^\sharp\|_F^2 \rangle,$$

which also shows that a sufficient condition for the $\{\tau_k\}$ to have square summable matrix 2-norm (in view of the fact that $\|A\|_2 \le \|A\|_F$ for any matrix A) is that $\|\gamma(z)^\sharp\|_F^2$ has finite integral. By the definition of adjoint,

$$\|\gamma(z)^\sharp\|_F^2 = \sum_{r=1}^N \sum_{s=1}^N |\det [f(\lambda)]_{s,r}|^2,$$

where the s,r subscript means that we remove the s-th row and r-th column before computing the determinant. Because $f(\lambda)$ is Hermitian, we do not need the absolute value bars, and for any r,s we obtain

$$\begin{aligned}
&|\det [f(\lambda)]_{s,r}|^2 = \det\left([f(\lambda)]_{s,r}[f(\lambda)]'_{s,r}\right) \le \left[\text{max eigenvalue}\left([f(\lambda)]_{s,r}[f(\lambda)]'_{s,r}\right)\right]^N \\
&= \|[f(\lambda)]_{s,r}\|_2^{2N} = \left\|\sum_{k\in\mathbb{Z}} e^{-i\lambda k}[\gamma_k]_{s,r}\right\|_2^{2N} \le \left(\sum_{k\in\mathbb{Z}} \|[\gamma_k]_{s,r}\|_2\right)^{2N} \\
&\le \left(\sum_{k\in\mathbb{Z}} \|[\gamma_k]_{s,r}\|_F\right)^{2N} \le \left(\sum_{k\in\mathbb{Z}} \|\gamma_k\|_F\right)^{2N}.
\end{aligned}$$

In the above, we have used the fact that for a non-negative definite matrix the determinant is bounded by the Nth power of the largest eigenvalue; we also used a well-known expression for the matrix 2-norm, took the sub-matrix of $f(\lambda)$ term by term, used the triangle inequality, noted that the 2-norm is bounded by the Frobenius norm, and finally used the fact that the Frobenius norm of a sub-matrix is bounded above by the Frobenius norm of the whole matrix. Because the matrix Frobenius norm is bounded by a constant times the matrix 2-norm, it follows that the Frobenius norm of the $\{\gamma_k\}$ is summable, and hence the $\{\tau_k\}$ have square summable matrix 2-norm.

Next, we show that $\det f(\lambda)$ has absolutely summable Fourier coefficients. From the identity

$$\det f(\lambda)\, I_N = f(\lambda)\, f(\lambda)^\sharp$$

we find that the Fourier coefficients of $\det f(\lambda)\, I_N$ are

$$\langle z^{-k} \det \gamma(z)\, I_N\rangle = \sum_{j\in\mathbb{Z}} \gamma_j\, \langle z^{j-k}\gamma(z)^\sharp\rangle.$$

Applying the triangle inequality and the Hölder inequality for sequences (and letting $C = \|I_N\|_2$),

$$\sum_{k\in\mathbb{Z}} |\langle z^{-k}\det\gamma(z)\rangle| \le C^{-1} \sqrt{\sum_{j\in\mathbb{Z}} \|\gamma_j\|_2^2}\, \sqrt{\sum_{j\in\mathbb{Z}} \|\langle z^{-j}\gamma(z)^\sharp\rangle\|_2^2}. \tag{A.3}$$

We have already shown square summability of the adjoint's Fourier coefficients (with respect to the matrix 2-norm), and the $\{\gamma_j\}$ have summable, and hence square summable, matrix 2-norm by Assumption B. Hence, $\det f(\lambda)$ has absolutely summable Fourier coefficients, and we can apply Wiener's lemma, thereby concluding that $1/\det f(\lambda)$ has an absolutely convergent Fourier series.

Finally, having shown the square summability of both factors on the right hand side of eq. (A.2), a Hölder inequality argument similar to that leading to eq. (A.3) implies that the Fourier coefficients of $f(\lambda)^{-1}$ have summable matrix 2-norm. This completes the proof, but we also show that any component of f or f^{-1} is absolutely integrable:

$$(2\pi)^{-1} \int_{-\pi}^{\pi} |[f(\lambda)]_{s,r}| d\lambda \leq \sum_{k\in\mathbb{Z}} |[\gamma_k]_{s,r}| \leq \sum_{k\in\mathbb{Z}} \|\gamma_k\|_F \leq \sqrt{N} \sum_{k\in\mathbb{Z}} \|\gamma_k\|_2,$$

which is finite for any $1 \leq r, s, \leq N$. A similar argument holds for f^{-1}. □