Appendices for “Locally stationary wavelet packet processes: basis selection and model fitting.”

Alessandro Cardinali∗ and Guy P. Nason†

November 18, 2016

Appendix

Proof of Theorem 1

The proof is based on three intermediate results: (i) Positive definiteness of $A$; (ii) Decays of entries for $A$; (iii) Boundedness of inverse $A^{-1}$.

(i) Positive definiteness of $A$. $A$ is positive semi-definite as it can be written as $A = \Psi'\Psi$ where $\Psi$ is a matrix constructed from autocorrelation wavelet packets and hence $x'Ax = x'\Psi'\Psi x = (\Psi x)'(\Psi x) = y'y \geq 0$ for some $y$. For definiteness, suppose $x'Ax = 0$ for some $x$. This implies $\sum_{\tau}\{\sum_p x_p \Psi_p(\tau)\}\{x_{p'}'\sum_{p'} \Psi_{p'}(\tau)\} = 0$. In other words $\sum_{\tau} a_{\tau}^2 = 0$ where $a_{\tau} = \sum_p x_p \Psi_p(\tau)$ and hence $a_{\tau} = 0$ for all $\tau$. Hence $\sum_p x_p \Psi_p(\tau) = 0$ for all $\tau$ and because of the linear independence of the $\{\Psi_p(\tau)\}_p$ we must have $x_p = 0$ for all $p$.

(ii) Decay of entries of $A$. We can use the alternative definition of $A$ in terms of cross-autocorrelation wavelets $\Psi_{p,p'}(\tau) = \sum_t \psi_p(t)\psi_{p'}(t+\tau)$:

$$A_{p,p'} = \sum_{\tau} \Psi_{p,p'}^2(\tau)$$

(1)

and express the decay of the entries of $A$ using those of $\Psi_{p,p'}(\tau)$. In the following we will use the two scale relationship $\psi_{j,i}(t) = \sum_k p_t \psi_{j-1,i'}(k)$, with $i' = \lfloor i/2 \rfloor$, the integer part (see Percival and Walden (2000), equation 231b). In the following we will use this relation in cascade. Note that this relation holds for any packet $(j,i)$, whether this is in a basis $b$ or not. Finally we will also use that

$$p_t^{j,i} = \begin{cases} h_t, & \text{if } i \mod 4 = 0, 3; \\ g_t, & \text{if } i \mod 4 = 1, 2. \end{cases}$$

(2)

and the definition of orthogonal wavelet packet filters $\sum_t p_t p_{t-2k} = \delta_k$ for all $(j,i)$. Since this relation holds for any packet, in the following we will simply write $p_t$. For $p, p' \in b$ and $p \neq p'$, let $j_p = j$ and $j_{p'} = l$. Moreover, we will also assume without loss of generality that there are at least two wavelet packets belonging to different scales $(j_p \neq j_{p'})$ such that $\min(j_p, j_{p'}) = \min(j, l) = l$. Setting $k - m = \tau$, by noting that as a cross-correlation (wavelet) function it has a peak at lag 0, we can bound using Cauchy-Schwartz as

$$\Psi_{p,p'}^2(\tau) \leq \Psi_{p,p}^2(0)\Psi_{p',p'}^2(0)$$

(3)

∗University of Plymouth. Email: alessandro.cardinali@plymouth.ac.uk

†University of Bristol. Email: g.p.nason@bristol.ac.uk
Indicating the wavelet packet filter’s transfer function as \( m_{u,j,b}(\omega), u = 0, 1 \) under the same conditions as (2), we have for \( \Psi_{p,p}^2(0) \)

\[
\Psi_{p,p}^2(0) = \left( \sum_k p_k \psi_j(k) \right)^2 = \int |\hat{p}(\omega)|^2 |\hat{\psi}_j(\omega)|^2 d\omega
\]

\[
= \int |m_u(\omega)|^2 \prod_{s=0}^j |m_{u,j,i}(2^s \omega)|^2 d\omega
\]

\[
\leq \int |m_0(\omega)|^2 \prod_{s=0}^j 2^{-s} |m_{0,j,i}(2^{-s} \omega)|^2 d\omega
\]

\[
\leq \int d\omega \prod_{s=0}^j 2^{-s} \leq 2 \pi K_1 \prod_{s=0}^j 2^{-s}
\]  

(4)

for some \( K_1 < \infty \). Analogously we have

\[
\Psi_{p',p'}^2(0) \leq 2 \pi K_2 \prod_{r=0}^l 2^{-r}
\]  

(5)

for some \( K_2 < \infty \). Using (3), (4) and (5) we have

\[
\Psi_{p',p'}^2(\tau) \leq K_3 \prod_{s=0}^j 2^{-s} \prod_{r=0}^l 2^{-r}
\]

\[
\leq K_3 \prod_{s=0}^j \prod_{r=0}^l 2^{-(s+r)}
\]  

(6)

for some positive \( K_3 < \infty \). For \( j > l \), using the difference of triangular numbers and completing the squares, we finally obtain

\[
\Psi_{p',p'}^2(\tau) \leq K_3 2^{- \sum_{s=0}^j \sum_{r=0}^l (s+r)}
\]

\[
\leq K_3 2^{- \sum_{s=0}^j \sum_{r=0}^l (s+r)} - \sum_{s=l+1}^j s
\]

\[
\leq K_3 2^{- \sum_{s=l+1}^j s}
\]

\[
\leq K_3 2^{-((j^2+j-l^2-l)/2)}
\]

\[
\leq K_3 2^{-((j-l)^2+2jl-2l^2+j-l)/2}
\]

\[
\leq K_3 2^{-(j-l)^2/2}
\]

\[
\leq K_3 2^{-l^2}
\]  

(7)

To obtain (4) we bounded the squared gain functions for a generic sequence with the squared gain for the scaling filter, using the properties of trigonometric polynomials. Hence the squared cross-correlation will have (lower bound) decay \( O(2^{-|j-l|}) \) and, in view of (1), \( A \) will have the same decay. As an illustration of the less general wavelet example see Figure 1, which plots numerically computed values of \( \log A_{1,j} \) versus the scale index \( j \), and show that the (known) ‘worst-case’ for decay is the Haar wavelet which ‘achieves’ the \( 2^{-j} \) decay rate of (7).

(iii) Boundedness of inverse.

**Definition 1 (GMRS-239)** A matrix \( A = (A_{i,j})_{i,j \in J} \) (where for our purposes the set \( J = \mathbb{Z} \) or \( \mathbb{Z}_+ \), the non-negative integers) decays exponentially if there exist constants \( c > 0 \) and \( 0 < \lambda < 1 \) such that \( |A_{i,j}| \leq c\lambda^{\left| i-j \right|} \) for all \( i, j \in J \).
Figure 1: Plot of numerically computed \( \log A_{1,j} \) for \( j = 1, \ldots, 20 \) for Haar wavelets plotted as \( \text{H} \), and Daubechies’ extremal phase wavelets with 4 and 8 vanishing moments. The slope of the Haar line obtained by linear regression is (numerically) exactly \(-\log 2\).

**Definition 2 (GMRS-240)** If \( A \) and \( B \) are semi-infinite matrices, we write \( A \sim B \) if there are constants \( c > 0 \) and \( 0 < \lambda < 1 \) such that \( |A_{ij} - B_{ij}| \leq c\lambda^i+j \) for all \( i,j \geq 0 \).

GMRS-240 note that if \( A,B \) are semi-infinite matrices such that \( B \) decays exponentially and \( A \sim B \) then \( A \) decays exponentially also. GMRS-235 guarantees that semi-infinite positive definite \( A \) can be written in Cholesky form as \( A = MM' \) where \( M \) is lower triangular with positive diagonal entries. Then, using the decay properties of \( A \) from (ii) above, construct a lower triangular \( L \) that decays exponentially such that \( A \sim LL' \). Then Theorem 2.2 of GMRS states that \( M \sim L \) and \( M \) has a lower triangular inverse \( M^{-1} \) with \( M^{-1} \sim L^{-1} \) and hence \( M^{-1} \) decays exponentially. From GMRS-Lemma 2.2 \( A^{-1} = (M^{-1})M^{-1} \) decays exponentially and from GMRS-Lemma 2.1 \( A^{-1} \) is bounded in \( l^2(\mathbb{Z}_+) \). This completes the proof.

**Proof of Proposition 1**

By omitting the constant term the (negative) Gaussian log-likelihood for the LSWP vector \( X_t \) is approximated by

\[
\mathcal{L}_T\{b, S(b)\} = \frac{1}{2T} \log |\Sigma_{S(b)}| + \frac{1}{2T} X'_T \Sigma_{S(b)}^{-1} X_T + O(T^{-1}),
\]

where \( \mathbb{E}X_tX'_t = \Sigma_{S(b)} + O(T^{-1}) \) is the covariance matrix for LSWP processes. This \( (T \times T) \) matrix can be decomposed as \( \Sigma_{S(b)} = V S V' \) where \( S = \text{diag}\{S_p(t/T)\}_{p,t} \) is a diagonal matrix of dimensions \((|b|T \times |b|T)\) containing all the spectral values. The rectangular block matrix \( V = (V_p)_p \) has dimensions \((T \times |b|T)\) and is formed by \(|b| \times |b| \) square matrices each of dimensions \((T \times T)\). Each row of the \( V_p \) blocks is made of circularly shifted NDWP vectors \( \Psi_{jp,ip} \) defined in Section 2 and is full-rank. This implies that \( V \) and \( V' \) are also full-rank and therefore their generalized inverses exist and will be denoted as \( V^{-1} \) and \((V')^{-1} \) respectively. We also define
the sample covariance matrix $\Sigma_{L(b)} = X_t X'_t$ that can be decomposed as $\Sigma_{L(b)} = V L V'$ where $V$ is defined as above and $L = \text{diag}(L_{p,t})_{p,t}$ has dimensions $(|b|T \times |b|T)$.

The discussion of Remark 7 (in the main paper) and references therein imply that this will be an asymptotically unbiased estimator for the covariance matrix $\Sigma_{S(b)}$ even if it will not exactly be symmetrical in finite samples. The first part of the log-likelihood (8) can be rewritten using the identity

$$\log |\Sigma_{S(b)}| = \prod_t v_{t,T},$$

(9)

where $v_{t,T} = \mathbb{E}(X_t - \hat{X}_t)^2$ is the prediction error variance and $\hat{X}_t$ is the best linear predictor for $X_t$, that can be obtained, given a basis $b \in \mathcal{B}$, in the same fashion as Fryzlewicz et al. (2003). In fact we have that in our case

$$v_{t,T} = \exp \left\{ \sum_{p \in b} \log S_p \left( \frac{t}{T} \right) \int |\psi_p(\omega)|^2 d\omega \right\}$$

$$= \exp \left\{ \sum_{p \in b} \log S_p \left( \frac{t}{T} \right) \right\},$$

(10)

since

$$\int |\psi_p(\omega)|^2 d\omega = \sum_k \psi^2_p(k) = 1$$

(11)

for all packets $p$. We therefore have

$$\log |\Sigma_{S(b)}| = \sum_t \sum_{p \in b} \log S_p \left( \frac{t}{T} \right).$$

(12)

The second part of the log-likelihood (8) can be rewritten using the properties of the trace operator since $X'_T (\Sigma_{S(b)}^{-1}) X_T = tr(X'_T \Sigma_{S(b)}^{-1} X_T)$, being scalars equivalent to their trace. Using the decompositions of covariance matrices we have

$$tr(X'_T \Sigma_{S(b)}^{-1} X_T) = tr(X'_T \Sigma_{L(b)}^{-1} X_T)$$

$$= tr(\Sigma_{L(b)} \Sigma_{S(b)}^{-1})$$

$$= tr(V L V' (V S V')^{-1})$$

$$= tr(V L V' (V')^{-1} S^{-1} V^{-1})$$

$$= tr(V L I S^{-1} V^{-1})$$

$$= tr(L S^{-1} V^{-1})$$

$$= tr(LS^{-1} I)$$

$$= \sum_{p \in b} \sum_t \frac{L_{p,t}}{S_p(t/T)}.$$

Using these two results the log-likelihood can be finally expressed as

$$\mathcal{L}_T \{ b, S(b) \} = \frac{1}{2T} \sum_{p \in b} \sum_t \log S_p(t/T) + \frac{1}{2T} \sum_{p \in b} \sum_t \frac{L_{p,t}}{S_p(t/T)} + O(T^{-1}),$$

(13)

which completes the proof.
Proof of Proposition 2

Equation (13) gives an explicit expression for the log-likelihood in terms of the spectral ordinates \( \{ S_p(t/T) \} \) which are the nuisance parameters in the basis selection process. To prove part 1, the profile likelihood is obtained by replacing the spectra in equation (13) with the unbiased periodogram to obtain

\[
\tilde{L}_T(b) = \frac{1}{2} \left( |b| + T^{-1} \sum_{p \in b} \sum_t \log L_{p,t} \right) + O(T^{-1}).
\]  

(14)

To prove part 2 we simply calculate the expectation of the log-likelihood given in equation (13) to obtain

\[
\mathbb{E} \tilde{L}_T(b, S(b)) = \frac{1}{2} \left| b \right| + T^{-1} \sum_{p \in b} \sum_t \log S_p(t/T) + O(T^{-1}).
\]

The bias is therefore given by

\[
\mathbb{E} \left[ \tilde{L}_T(b) - \tilde{L}_T(b, S(b)) \right] = \frac{1}{2T} \sum_{p \in b} \sum_t \mathbb{E} \left[ \log L_p(t/T) - \log S_p(t/T) \right] + O(T^{-1}),
\]

and the result follows from the asymptotic unbiasedness of \( L_{p,t} \) and the application of Jensen’s inequality.

References

Fryzlewicz P, Van Bellegem S, von Sachs R. 2003. Forecasting non-stationary time series by wavelet process modelling. *Annals of the Institute of Statistical Mathematics* 55: 737–764.

Percival DB, Walden AT. 2000. *Wavelet Methods for Time Series Analysis*. Cambridge University Press.