Reconciling the Gaussian and Whittle Likelihood with an application to estimation in the frequency domain

CIRM, Luminy, September, 2020

Suhasini Subba Rao

Texas A&M University

Joint work with Junho Yang (final graduate student at Texas A&M).

Parameter estimation methods

- In general, there are two different approaches for fitting second order stationary models to a time series.

- Time Domain approach Construction is based on the Gaussian likelihood. The idea is to "fit" a parametric autocovariance of the form $\{c_\theta(r)\}_{r\in\mathbb{Z}}$ to the time series, where θ is the unknown parameter.

- Frequency Domain approach Construction is based on the Whittle likelihood. We "fit" a parametric spectral density form $f_\theta(\omega)$ to the periodogram of the time series.

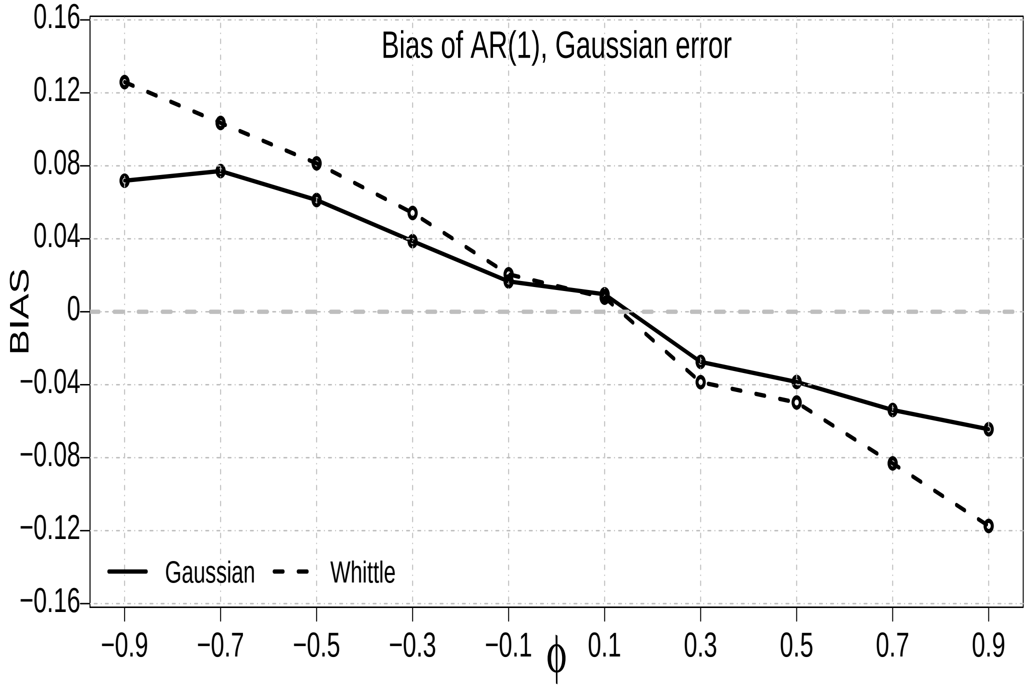
- The models are equivalent, as the autocovariance and spectral density are connected through the relation

$$f_\theta(\omega) = \sum_{r\in\mathbb{Z}} c_\theta(r)\exp(ir\omega).$$

- This relation connects the two parameterisations, but not the two estimation approaches.

Motivation: Bias of Gaussian and Whittle estimators

AR(1): $X_t = \phi X_{t-1} + \varepsilon_t, \ |\phi| < 1$ sample size $n = 20$.



- The difference in the bias is $\phi/n + O(n^{-3/2})$.
- In general, the difference is due to the derivative of the difference between the likelihoods.

Objective: Quantify the difference between the likelihoods.

Overview

- Using ideas from linear prediction and biorthogonal transforms, we show that the Gaussian likelihood has a frequency domain representation.

- Using this we connect the Whittle and Gaussian likelihood (circulant and inverse Toeplitz matrices) through a series expansion.

  The construction hinges on the "so called" predictive DFT.

- <u>Applications: Theoretical</u> Apply these results for obtaining an interpretable bound between the Gaussian and Whittle likelihoods, their derivatives and bias.

- <u>Applications: Practical</u> Obtain a new frequency domain criterion that combines the benefits of the Gaussian and Whittle likelihoods.

Definition: The Gaussian likelihood

- We observe the time series $\underline{X}_n = (X_1, \ldots, X_n)'$. Fit autocovariance $\{c_\theta(r)\}$ to the time series.

- The Gaussian Likelihood

$$\mathcal{L}_n(\theta; \underline{X}_n) = n^{-1}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n + n^{-1}\log|\Gamma_n(f_\theta)|$$

  $\Gamma_n(f_\theta)$ the variance matrix of $\underline{X}_n$.

- Due to Stationarity, $\Gamma_n(f_\theta)$ is an $n \times n$ Toeplitz matrix:

$$\Gamma_n(f_\theta) = \begin{pmatrix} c_\theta(0) & c_\theta(1) & \ldots & c_\theta(n-1) \\ c_\theta(1) & c_\theta(0) & \ldots & c_\theta(n-2) \\ c_\theta(2) & c_\theta(1) & \ldots & c_\theta(n-3) \\ \vdots & \vdots & \ddots & \vdots \\ c_\theta(n-1) & c_\theta(n-2) & \ldots & c_\theta(0) \end{pmatrix}$$

Definition: The Whittle likelihood

- The Whittle Likelihood (based on the Riemann sum). Fit the spectral density $f_\theta(\cdot)$ to the periodogram of the data.

$$K_n(\theta; \underline{X}_n) = n^{-1}\sum_{k=1}^{n}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} + n^{-1}\sum_{k=1}^{n}\log f_\theta(\omega_{k,n})$$

- It is based on the DFT (linear transform of the time series)

$$J_n(\omega_{k,n}) = n^{-1/2}\sum_{t=1}^{n} X_t \exp(it\omega_{k,n}).$$

with $[0, 2\pi]$ divided into a grid $\omega_{k,n} = 2\pi k/n$.

The Whittle likelihood in matrix form

- To compare the two likelihoods we rewrite the Whittle likelihood in matrix form

$$n^{-1}\sum_{k=1}^{n}\frac{|J_n(\omega_{k,n})|^2}{f_\theta(\omega_{k,n})} = n^{-1}\underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) F_n \underline{X}_n,$$

where

$$\Delta_n(f_\theta^{-1}) = \begin{pmatrix} f_\theta(\omega_{1,n})^{-1} & 0 & \ldots & 0 \\ 0 & f_\theta(\omega_{2,n})^{-1} & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ 0 & 0 & \ddots & f_\theta(\omega_{n,n})^{-1} \end{pmatrix},$$

- F_n is the discrete Fourier transform matrix, $(F_n)_{k,t} = n^{-1/2}\exp(it\omega_{k,n})$.
  $(F_n\underline{X}_n)_k = J_n(\omega_{k,n}) = n^{-1/2}\sum_{t=1}^{n} X_t \exp(it\omega_{k,n})$.

Comparing the matrices in the two likelihoods

- Simple calculations show that the Whittle matrix $F_n^*\Delta_n(f_\theta^{-1})F_n$ is circulant

$$F_n^*\Delta_n(f_\theta^{-1})F_n = \begin{pmatrix} a_\theta(0) & a_\theta(1) & \ldots & a_\theta(n-1) \\ a_\theta(n-1) & a_\theta(0) & \ldots & a_\theta(n-2) \\ a_\theta(n-2) & a_\theta(n-1) & \ldots & a_\theta(n-3) \\ \vdots & \vdots & \ddots & \vdots \\ a_\theta(1) & a_\theta(2) & \ldots & a_\theta(0) \end{pmatrix},$$

where $a_\theta(0)$ is the wrapped inverse covariance (corresponding to f_θ^{-1}) and all rows are cyclic permutations of the first row.

- On the other hand, for the Gaussian likelihood $\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n$, the inverse Toeplitz $\Gamma_n(\theta)^{-1}$ is neither circulant of Toeplitz.

Example: AR(1)

- The inverse corresponding to the AR(1) model $X_t = \phi X_{t-1} + \varepsilon_t$

$$
\Gamma_n(f_\theta)^{-1} = \begin{pmatrix} 1 & -\phi & 0 & 0 & \dots & 0 \\ -\phi & 1+\phi^2 & -\phi & 0 & \dots & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix}
$$

$$
F_n^* \Delta_n(f_\theta^{-1}) F_n = \begin{pmatrix} 1+\phi^2 & -\phi & 0 & 0 & \dots & -\phi \\ -\phi & 1+\phi^2 & \phi & 0 & \dots & 0 \\ 0 & -\phi & 1+\phi^2 & -\phi & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -\phi & 0 & 0 & 0 & \dots & 1+\phi^2 \end{pmatrix}
$$

- For AR(1) the difference lies only at the four corners of the matrix.

Known bounds between the Gaussian and Whittle

- If $\sum_r |c_\theta(r)| < \infty$, it is well known (see eg. Dahlhaus (1988)), then $\|\Gamma_n(f_\theta)^{-1} - F_n^* \Delta_n(f_\theta^{-1}) F_n\|_2 = o(1)$.

  Based on this, the Whittle likelihood is an approximation of the Gaussian likelihood. Leading to the well known result

$$\underline{X}_n' \Gamma_n(f_\theta)^{-1} \underline{X}_n \approx \underline{X}_n' F_n^* \Delta_n(f_\theta^{-1}) F_n \underline{X}_n.$$

- In the next few slides we obtain a construction to explicitly connects $F_n^* \Delta_n(f_\theta^{-1}) F_n$ with $\Gamma_n(f_\theta)^{-1}$.

First step: Biorthogonal transforms

- Let U_n and V_n denote two "transformation" matrices.

  $U_n\underline{X}_n$ and $V_n\underline{X}_n$ may not be orthogonal, in the sense that $\mathrm{var}(U_n\underline{X}_n) \neq$ diagonal and $\mathrm{var}(V_n\underline{X}_n) \neq$ diagonal.

- $U_n\underline{X}_n$ and $V_n\underline{X}_n$ are said to be biorthogonal w.r.t. $\Gamma_n = \mathrm{var}(\underline{X}_n)$ if

$$\mathrm{cov}\left(U_n\underline{X}_n, V_n\underline{X}_n\right) = U_n\Gamma_n V_n^* = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$$

- Application: Matrix inversion identity

$$\Gamma_n^{-1} = V_n^*\mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1})U_n,$$

- Application: Quadratic forms w.r.t. Γ_n^{-1}:

$$\begin{aligned} \text{since } \operatorname{cov}\left([U_n \underline{X}_n]_k, [V_n \underline{X}_n]_k\right) &= \lambda_k \\ \Rightarrow \underline{X}_n' \Gamma_n^{-1} \underline{X}_n &= \sum_{k=1}^{n} \frac{[U_n \underline{X}_n]_k [V_n^* \underline{X}_n]_k}{\lambda_k} \end{aligned}$$

where $[\underline{a}]_k$ denotes the kth entry of the vector $\underline{a}$.

- This is a generalisation of the spectral decomposition spectral decomposition:

$$\Gamma_n^{-1} = E_n^* \mathrm{diag}(\lambda_1^{-1}, \ldots, \lambda_n^{-1}) E_n.$$

Application to the Whittle and Gaussian likelihood

Gaussian $\Gamma_n(f_\theta)^{-1}$ Whittle $F_n^* \Delta_n(f_\theta^{-1}) F_n$.

- Recall the kth entry of $F_n \underline{X}_n$ is $J_n(\omega_{k,n})$.
- Objective Find the transformation, $U_n \underline{X}_n$, that is biorthogonal to $F_n \underline{X}_n$.
- This will allow is to invert Γ_n^{-1} in terms of F_n and U_n.

  We will show that the diagonal matrix is $\mathrm{diag}(f_\theta(\omega_{1,n}), \ldots, f_\theta(\omega_{1,n}))$.
- Next Step Obtain U_n.

Starting point

- It is well known that DFTs of a second order stationary time series with $\sum_r |rc(r)| < \infty$, are "almost" uncorrelated:

$$\text{cov}\left(J_n(\omega_{k_1,n}), J_n(\omega_{k_2,n})\right) = f(\omega_{k_1,n})\delta_{k_1,k_2} + O(n^{-1})$$

  where δ_{k_1,k_2} is the indicator variable.

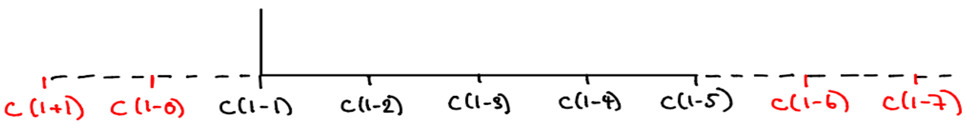
- Where does the above come from, and why $O(n^{-1})$?

- The derivation is based on the elementary sum of exponentials identity

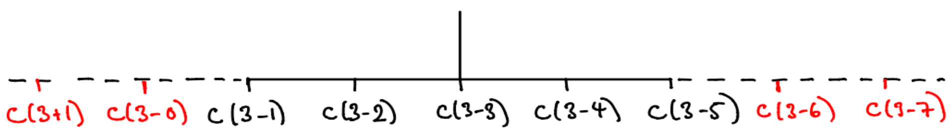
$$\frac{1}{n}\sum_{t=1}^{n} \exp\left(it\omega_{k_1-k_2,n}\right) = \begin{cases} 0 & k_1 - k_2 \notin n\mathbb{Z} \\ 1 & k_1 - k_2 \in n\mathbb{Z} \end{cases}.$$

$$\mathrm{cov}\left[J_n(\omega_{k_1,n}), J_n(\omega_{k_2,n})\right] = \frac{1}{n}\sum_{t=1}^{n} e^{it(\omega_{k_1,n}-\omega_{k_2,n})}\sum_{\tau=1}^{n} c(t-\tau)e^{i(t-\tau)\omega_{k_2,n}}$$

change to $$\frac{1}{n}\sum_{t=1}^{n} e^{it(\omega_{k_1,n}-\omega_{k_2,n})}\sum_{\tau=-\infty}^{\infty} c(t-\tau)e^{i(t-\tau)\omega_{k_2,n}} = f(\omega_{k_1,n})\delta_{k_1,k_2}$$



$$\sum_{t=1}^{5} c(1-t)e^{i(1-t)\omega_{k,n}}$$



$$\sum_{t=1}^{5} c(3-t)e^{i(3-t)\omega_{k,n}}$$

Aim Ammend one of the DFTs, by finding random variables that will reproduce the red covariances and allow us to extend the boundary, without the $O(n^{-1})$ error.

Extending the boundary

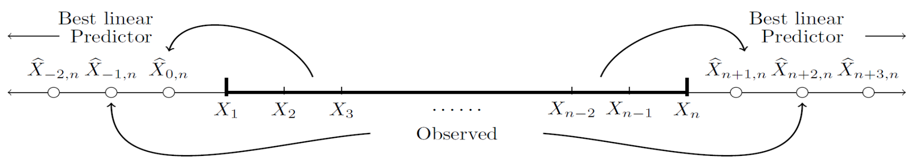
- Apply results from linear prediction Let $\mathbf{X} = \mathrm{sp}(X_1, \ldots, X_n)$ and $P_{\mathbf{X}}(Y)$ denote the linear projection of the random variable Y onto $\mathbf{X}$. We use the well know result

$$\mathrm{cov}\left[P_{\mathbf{X}}(Y), X_\ell\right] = \mathrm{cov}(Y, X_\ell) \qquad 1 \leq \ell \leq n.$$

- Set $Y = X_\tau$ for $\tau \neq \{1, \ldots, n\}$. Then we have

$$\mathrm{cov}\left[P_{\mathbf{X}}(X_\tau), X_t\right] = c(\tau - t) \qquad 1 \leq t \leq n.$$

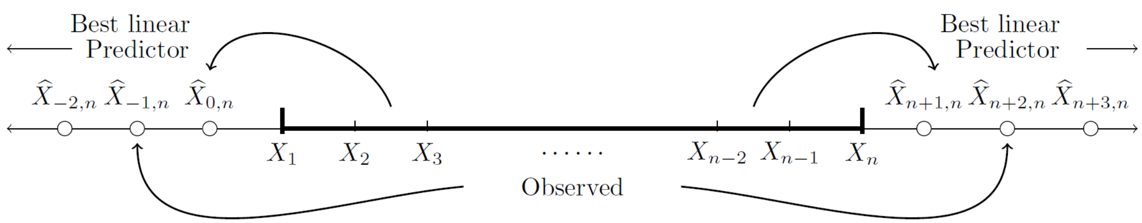
- Define $\widehat{X}_{\tau,n} = P_{\mathbf{X}}(X_\tau)$. Replace one DFT with $\sum_{\tau=-\infty}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega_{k,n}}$



Theorem Suppose $\sum_r |c(r)| < \infty$. Let

$$\underbrace{\widetilde{J}_n(\omega; f)}_{\text{Complete DFT}} = n^{-1/2} \sum_{\tau=-\infty}^{\infty} \widehat{X}_{\tau,n} e^{i\tau\omega_{k,n}} = J_n(\omega) + \underbrace{\widehat{J}_n(\omega; f)}_{\text{predictive DFT}}$$

$$\text{where } \widehat{J}_n(\omega; f) = n^{-1/2} \sum_{\tau \leq 0} \widehat{X}_{\tau,n} e^{i\tau\omega} + n^{-1/2} \sum_{\tau > n} \widehat{X}_{\tau,n} e^{i\tau\omega};$$



Then

$$\text{cov}\left[\widetilde{J}_n(\omega_{k_1,n}; f), J_n(\omega_{k_2,n})\right] = f(\omega_{k_1,n})\delta_{k_1,k_2}.$$

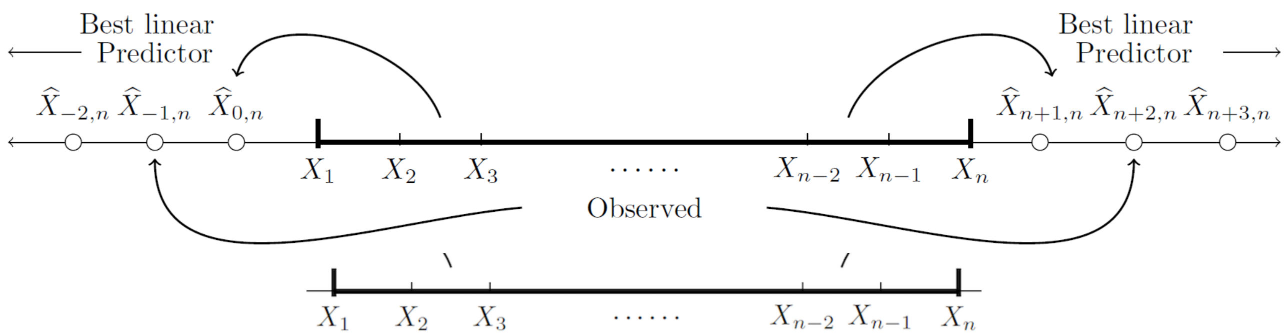
The biorthogonal transform of the DFT

- Interpretation Since $\widetilde{J}_n(\omega_{k,n};f) \in \mathrm{sp}(X_1,\ldots,X_n)$, then

$$\{\widetilde{J}_n(\omega_{k,n};f)\}_{k=1}^{n} \overset{\text{Biorthogonal}}{\Longleftrightarrow} \{J_n(\omega_{k,n})\}_{k=1}^{n},$$

where $\widetilde{J}_n(\omega_{k,n};f)$ is the regular DFT together with an additional term which predicts across the boundary of observation.

With $\mathrm{cov}[\widetilde{J}_n(\omega_{k,n};f), J_n(\omega_{k,n})] = f(\omega_{k,n})$.



Frequency representation of the Gaussian likelihood

Theorem If $\sum_r |c_\theta(r)| < \infty$, then (ignoring the $\log$ term) we have the frequency domain representation

$$\begin{aligned}
\mathcal{L}_n(\theta) &= \frac{1}{n}\underline{X}_n'\Gamma_n(f_\theta)^{-1}\underline{X}_n \\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{[U_n\underline{X}_n]_k[F_n^*\underline{X}_n]_k}{\lambda_k} \\
&= \frac{1}{n}\sum_{k=1}^{n}\frac{\widetilde{J}_n(\omega_{k,n};f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}
\end{aligned}$$

- Observe If we *fit* f_θ to the data, then the linear predictors in the Gaussian likelihood are based on the autocovariance associated with f_θ.

Difference between the two likelihoods

$$\text{Difference} \quad = \quad \text{Gaussian} - \text{Whittle} = \frac{1}{n}\sum_{k=1}^{n} \frac{\widehat{J}_n(\omega_{k,n}; f_\theta)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}$$

- The Gaussian likelihood simultaneously fits f_θ and predicts. The Whittle likelihood only fits $f_\theta(\cdot)$ to $|J_n(\cdot)|^2$.

- The Whittle likelihood is biased due to the "hard" truncation at the boundary of observation (outside $X_1, \ldots, X_n$).

- Next, our focus will on the "predictive DFT" $\widehat{J}_n(\omega_{k,n}; f_\theta)$.

Expression for the predictive DFT

- Supose the coeffiicents $\{\phi_{t,n}(\tau;f)\}_{t=1}^{n}$ minimize the L_2-distance $\mathbf{E}_f[X_\tau - \sum_{t=1}^{n}\phi_{t,n}(\tau;f)X_t]^2$. Then the best linear predictor of X_τ given $\underline{X}_n$ is

$$\widehat{X}_{\tau,n} = \sum_{t=1}^{n}\phi_{t,n}(\tau;f)X_t.$$

- The predictive DFT is

$$\widehat{J}_n(\omega;f) = n^{-1/2}\sum_{t=1}^{n}X_t\sum_{\tau\leq 0}\phi_{t,n}(\tau;f)e^{i\tau\omega} + \text{reflective term}.$$

Example: Predictive DFT of the AR(1) model

- Suppose f corresponds to the AR(1) model: $X_t = \phi X_{t-1} + \varepsilon_t$.

  It yields the predictors $\widehat{X}_0 = \phi X_1$, $\widehat{X}_{-1} = \phi^2 X_1$, ...on the left

  and $\widehat{X}_{n+1} = \phi X_n$, $\widehat{X}_{n+1} = \phi^2 X_n$,... on the right.

$$\widehat{J}_n(\omega; f_\theta) = \frac{\phi}{\sqrt{n}} \left(\frac{1}{\phi(\omega)} X_1 + \frac{e^{i(n+1)\omega}}{\overline{\phi(\omega)}} X_n \right) \qquad \phi(\omega) = 1 - \phi e^{-i\omega}.$$

Predictive DFT of AR(p) models

- $f_\theta = \sigma^2 |1 - \sum_{j=1}^{p} \phi_j X_{t-j}|^{-2} = \sigma^2 |\phi_p(\omega)|^{-2}$, where $p < n$.

- The predictive DFT has an analytic form in terms of $\underline{\phi}$

$$\widehat{J}_n(\omega; f_\theta) = \frac{n^{-1/2}}{\phi_p(\omega)} \sum_{t=1}^{p} X_t \underbrace{\sum_{s=0}^{p-t} \phi_{t+s} e^{-is\omega}}_{=\zeta_{t,n}^{(1)}(\omega;\phi)} + \text{reflective term}$$

- $\text{Difference} = \text{Gaussian} - \text{Whittle}$

$$= \frac{\sigma^{-2}}{n} \sum_{\ell=1}^{p} X_{n+1-\ell} \sum_{s=0}^{p-\ell} \phi_{\ell+s} \left(X_{(s+1) \bmod n} - \sum_{j=1}^{p} \phi_j X_{(s+1-j) \bmod n} \right) +$$

The predictive DFT for general spectral densities

- We showed that for AR(p) models where $n > p$, that $\widehat{J}_n(\omega; f)$ has an analytic form.

- For general spectral density functions, $\widehat{J}_n(\omega; f)$ does not have a simple analytic form.

- Suppose $f(\omega)$ which are bounded away from zero and from above, then by using Szegö (1921) and Baxter (1962) we have the decomposition

  $f(\omega) = \sigma^2 |1 - \sum_{j=1}^{\infty} \phi_j e^{ij\omega}|^{-2} = \sigma^2 |\phi(\omega)|^{-2}$, where $\{\phi_j\}$ are causal/minimum phase coefficients (this is the AR(∞) representation of a general time series).

- We obtain a series expansion of $\widehat{J}_n(\omega; f)$ in terms the AR(∞) coefficients.

- Series expansion By using von Neumann's alternating projections theorem and Inoue and Kasahara (2006) we can represent the finite predictions $\phi_{t,n}(\tau)$ in terms of AR and MA coefficients. We use this to obtain the decomposition

$$\widehat{J}_n(\omega; f) = \sum_{s=1}^{\infty} \widehat{J}^{(s)}_{\infty,n}(\omega; \phi)$$

- Each term $\widehat{J}^{(s)}_{\infty,n}(\omega; \phi)$ can be written as

$$\widehat{J}^{(s)}_n(\omega; \phi) = \frac{n^{-1/2}}{\phi(\omega)} \sum_{t=1}^{n} X_t \zeta^{(s)}_{t,n}(\omega; \phi)$$

where $\zeta^{(s)}_{t,n}(\omega; \phi)$ is a recursive integral in terms of the AR(∞) coefficients.

- The first term in the series expansion is

$$\widehat{J}_n^{(1)}(\omega;\phi) \quad = \quad \frac{n^{-1/2}}{\phi(\omega)} \sum_{t=1}^{n} X_t \sum_{s=0}^{\infty} \phi_{t+s} e^{-is\omega} + \text{reflective term}.$$

  This is a generalisation of the AR(p) result to AR(∞).

- Using Baxter-type inequalities we can show that

$$\widehat{J}_n(\omega;f) = \widehat{J}_n^{(1)}(\omega;\phi) + O\left(\frac{1}{n^{K-1/2}}\right)$$

  where $\sum_r |r^K c(r)| < \infty$.

Application: Theoretical bounds

The series expansion and approximation for $\widehat{J}_n(\omega; f)$ is used to analyze

- $\Gamma_n(f_\theta)^{-1} - F_n^* \Delta_n(f_\theta^{-1}) F_n$

- Difference between the likelihoods $\mathcal{L}_n(\theta) - K_n(\theta)$

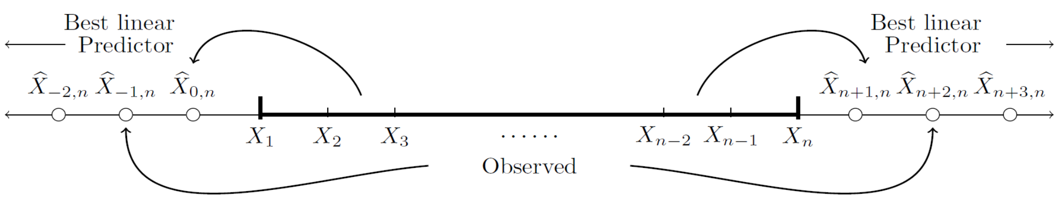
- Difference between derivatives of the likelihoods $\nabla_\theta^\ell[\mathcal{L}_n(\theta) - K_n(\theta)]$.

- Difference between asymptotic bias of Whittle and Gaussian likelihood estimators.

Application: Estimation

Next objective:

- Estimate the predictive DFT from data.



- Develop a frequency domain criterion which is a hybrid of the Whittle and the Gaussian likelihood.

- Potential benefits: it has the computational simplicity of the Whittle likelihood but tends to have the performance of the Gaussian likelihood.

Estimating the predictive DFT

- We have shown

$$\widehat{J}_n(\omega;\phi) = \frac{n^{-1/2}}{\phi(\omega)} \sum_{t=1}^{n} X_t \sum_{s=0}^{\infty} \phi_{t+s} e^{-is\omega} + O\left(\frac{1}{n^{K-1/2}}\right).$$

- Fact The AR(∞) parameters can be approximated with the best fitting AR(p) parameters (see Baxter (1962) and Kreiss, Paparoditis and Politis (2011)).

- Idea Replace the AR(∞) parameters in $\widehat{J}_n^{(1)}(\omega_{k,n};\phi)$ with the best fitting AR(p) parameters (the plug-in estimator: Bhansali (1996) and Kley, Preuss and Fryzlewicz (2019)).

The plug-in estimator

- Given the time series $\{X_t\}$ select the order p using the AIC.

- Estimate the best fitting AR(p) parameters with the Yule-Walker estimator to give $\widehat{f}_p = \widehat{\sigma}^2|1 - \sum_{j=1}^{p} \widehat{\phi}_j e^{ij\omega}|^{-2}$.

- Replace $\widehat{J}_n(\omega_{k,n}; f) \Rightarrow \widehat{J}_n(\omega_{k,n}; \widehat{f_p})$ where

$$\widehat{J}_n(\omega; \widehat{f}_p) = \frac{n^{-1/2}}{\widehat{\phi}_p(\omega)} \sum_{t=1}^{p} X_t \sum_{s=0}^{p-t} \widehat{\phi}_{t+s} e^{-is\omega} + \text{ reflective term}$$

New likelihoods: The spectral divergence

$$\text{Spectral Divergence } I_n(f; f_\theta) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{f(\omega_{k,n})}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n})\right).$$

It is a measure of "distance" between the f and f_θ, which is smallest when $\theta_0 = \arg\min_\theta I_n(f; f_\theta)$.

Under correct specification $f = f_{\theta_0}$ The Gaussian likelihood $\mathbf{E}_{f_{\theta_0}}[\mathcal{L}_n(\theta_0)] = I(f_{\theta_0}; f_{\theta_0})$.

Under misspecification $f \neq f_{\theta_0}$ The Gaussian likelihood is $\mathbf{E}_f[\mathcal{L}_n(\theta_0)] = I(f; f_{\theta_0}) + O(n^{-1})$.

A criterion based on the spectral divergence

- If the true density is f, then predicting with the true density gives $\mathbf{E}_f[\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}] = f(\omega_{k,n})$.

- Based on this, define the infeasible criterion

$$
\begin{aligned}
L_n(\theta) &= \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\widetilde{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n})\right) \\
&= \text{Whittle} + \frac{1}{n}\sum_{k=1}^{n}\frac{\widehat{J}_n(\omega_{k,n}; f)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})}
\end{aligned}
$$

  Clearly $\mathbf{E}[L_n(\theta)] = I_n(f; f_\theta)$.

- Question Can we estimate L_n and still improve on $O(n^{-1})$?

A feasible criterion

$$\widehat{L}_{p,n}(\theta) \quad = \quad \frac{1}{n}\sum_{k=1}^{n}\left[\frac{\widetilde{J}_n(\omega_{k,n};\widehat{f}_p)\overline{J_n(\omega_{k,n})}}{f_\theta(\omega_{k,n})} + \log f_\theta(\omega_{k,n})\right].$$

Theorem Suppose $\sum_{r\in\mathbb{Z}}|r^K c(r)| < \infty$ and under suitable regularity conditions

$$\widehat{L}_{p,n}(\theta) = L_n(\theta) + O\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right)$$

and

$$|\widehat{\theta}_n - \widetilde{\theta}_n|_1 = O_p\left(\frac{p^3}{n^{3/2}} + \frac{1}{np^{K-1}}\right),$$

where $\widehat{\theta}_n = \arg\min \widehat{L}_{p,n}(\theta)$ and $\widetilde{\theta}_n = \arg\min L_n(\theta)$.

Sampling properties of new likelihood estimators

- Since the feasible and infeasible estimators are asymptotically equivalent in the sense $n|\widehat{\theta}_n - \widetilde{\theta}_n|_1 = o(1)$, the asymptotic sampling properties of the infeasible estimator holds for the feasible estimator.

- Under certain regularity conditions we obtain an expression for the asymptotic bias of $\widehat{\theta}_n$.

- The asymptotic variance of $\widehat{\theta}_n$ is equivalent to the asymptotic variance of the Whittle and Gaussian likelihood estimators.

Simulations

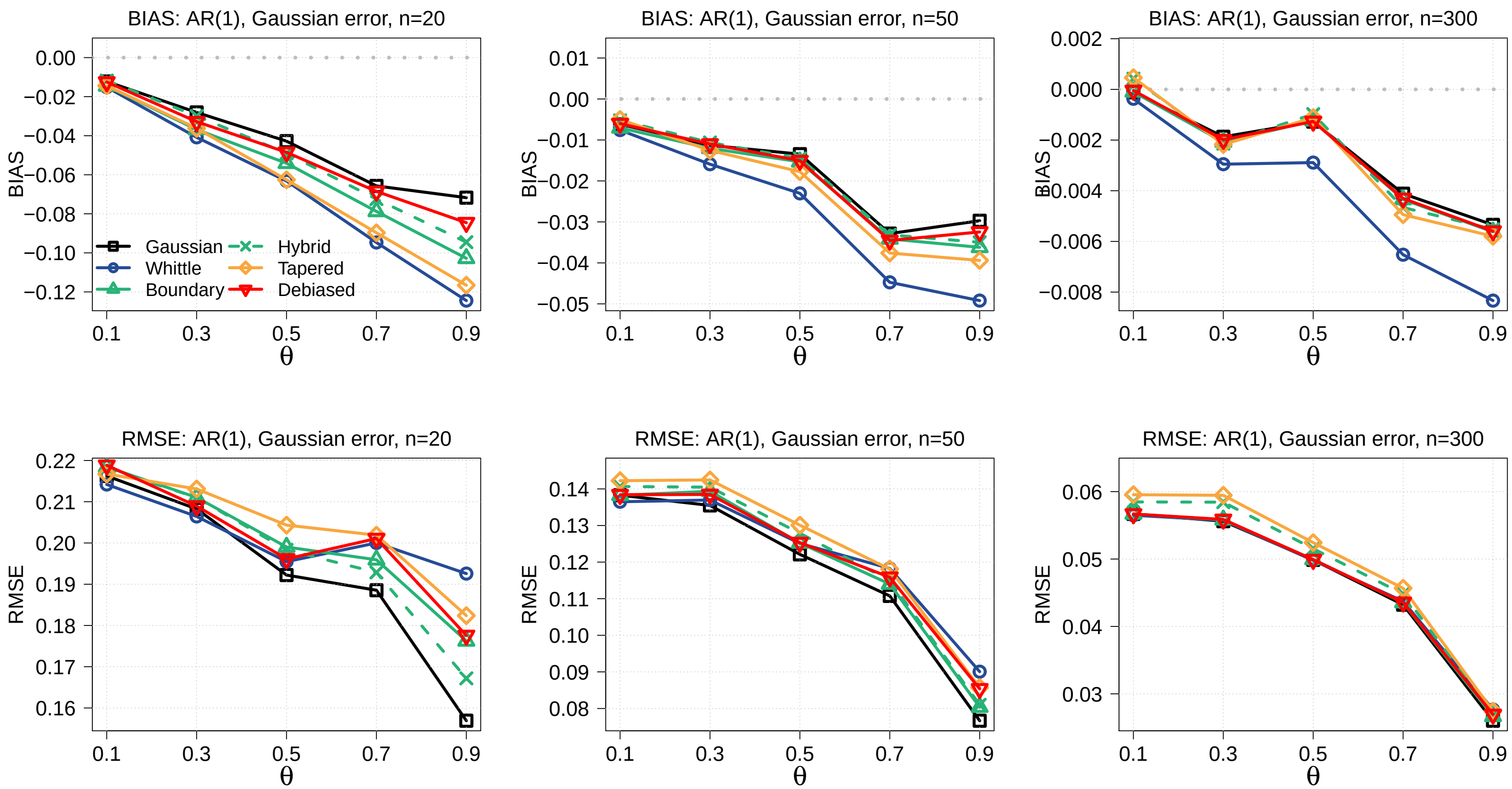
- We compare the new likelihood (two variants) with the Gaussian, Whittle, taper Whittle and debiased Whittle (proposed in Sykulski et. al. (2019)).

- **Model 1** We use the AR(1) model $X_t = \theta X_{t-1} + \varepsilon_t$ ($\theta = 0.1, 0.3, 0.5, 0.7, 0.9$) as the data generating process and sample sizes $n = 20$, 50 and 300.

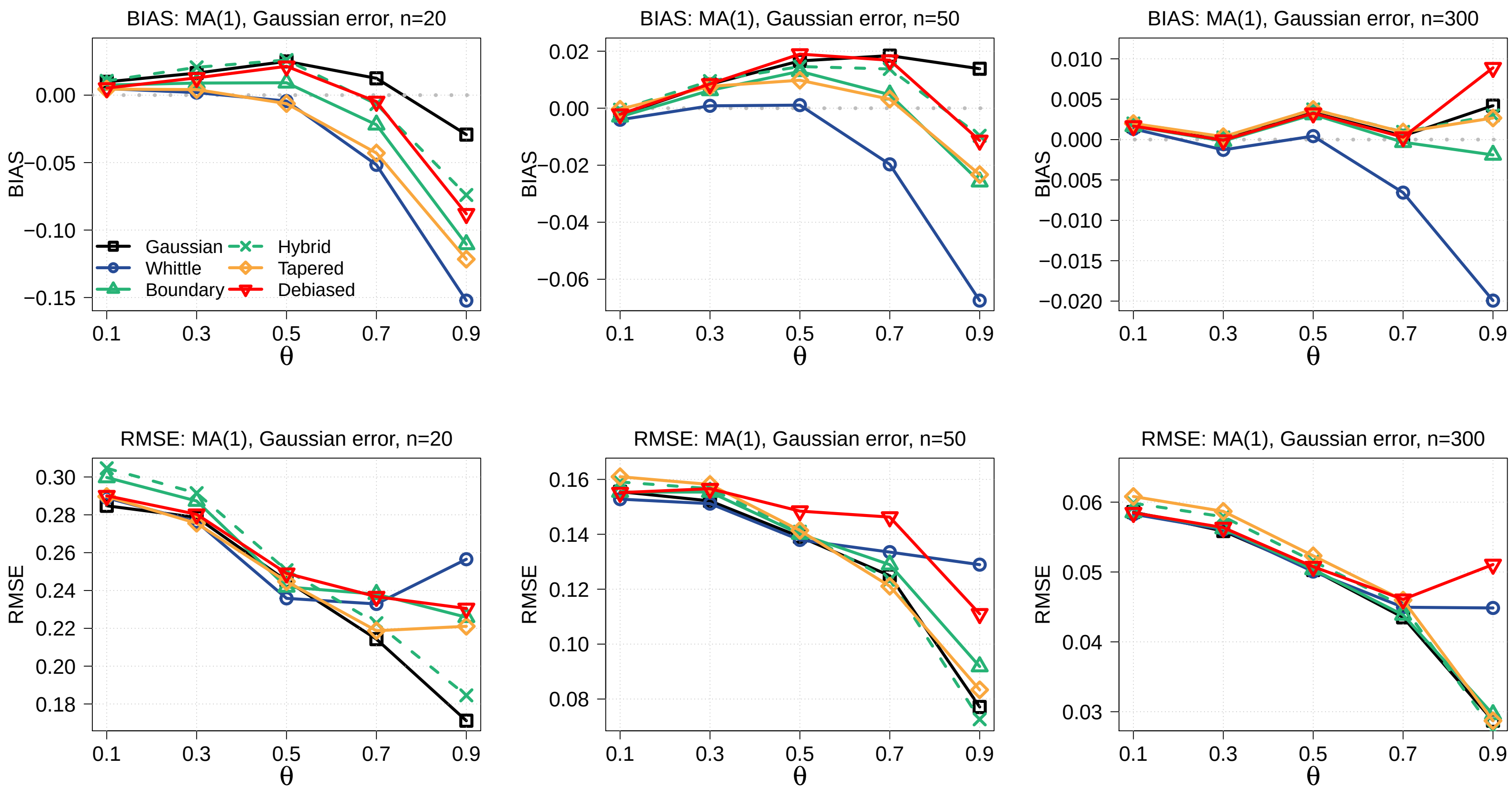
- **Model 2** We use the MA(1) model $X_t = \varepsilon_t + \theta\varepsilon_t$ ($\theta = 0.1, 0.3, 0.5, 0.7, 0.9$) as the data generating process and sample sizes $n = 20$, 50 and 300.

- Simulations under misspecification (but not presented here).

AR(1) specified



$\mathbf{MA}(1)$ specified



Current and future work

- Apply the "complete" periodogram $\widetilde{J}_n(\omega_{k,n}; \widehat{f}_p)\overline{J_n(\omega_{k,n})}$ to spectral density estimation.

- Generalisation of these results to the multivariate-time series framework.

- All results described in this talk are for short memory time series. We are currently studying the case of long memory time series.