Parametric or nonparametric: the FIC approach for stationary time series

Gudmund Horn Hermansen*
Department of Mathematics, University of Oslo, Oslo, Norway - gudmunhh@math.uio.no

Nils Lid Hjort
Department of Mathematics, University of Oslo, Oslo, Norway - nils@math.uio.no
Martin Jullum
Department of Mathematics, University of Oslo, Oslo, Norway - martinju@math.uio.no

Abstract

We seek to narrow the gap between parametric and nonparametric modelling of stationary time series processes. The approach is inspired by recent advances in focused inference and model selection techniques.Our paper generalises and extends current work by developing a new version of the focused information criterion (FIC), directly comparing the performance of both parametric and nonparametric time series models. This is achieved by comparing the mean squared error for estimating a focus parameter under consideration, for each candidate model. In particular this yields FIC formulae for covariances or correlations at specified lags, for the probability of reaching a threshold, etc. Suitable weighted average versions, the AFIC, also lead to model selection strategies for finding the best model for the purpose of estimating e.g. a sequence of correlations.
Keywords: focused inference; model selection; time series modelling; risk estimation.

1. Introduction and summary

The focused information criterion (FIC) was introduced in Claeskens & Hjort (2003) and is based on estimating and comparing the accuracy of individual model-based estimators for a chosen focus parameter. This focus, say μ, ought to have a clear statistical interpretation across candidate models. For a given candidate model, μ is then expressed as a function of this model's parameters. In general, the focus parameter can be any sufficiently smooth and regular function of the underlying model parameters, e.g. quantiles, regression coefficients, a specified lagged correlation and various types of predictions and data dependent functions, to name some; see Hermansen & Hjort (2015) for a more complete list and discussion of valid foci for time series models.

Suppose there are candidate models $M_1, \ldots, M_k$, leading to focus parameter estimates $\widehat{\mu}_1, \ldots, \widehat{\mu}_k$, respectively. The underlying idea leading to the FIC is to estimate the mean squared error (mse) of $\widehat{\mu}_j$ for each candidate model and then select the model that achieves the smallest value. The mse in question is $r_j = \mathrm{E}\,(\widehat{\mu}_j - \mu_{\rm true})^2 = \mathrm{Var}\,\widehat{\mu}_j + \mathrm{bias}(\widehat{\mu}_j)^2$, comprising the variance and the squared bias in relation to the true parameter value $\mu_{\rm true}$. Thus the FIC consists of finding ways of assessing, approximating and then estimating the r_j for each candidate model, and the winning model is the one with smallest $\widehat{r}_j$. How this may be done depends on both the candidate models and the focus parameter, as well as on other characteristics of the underlying situation. The FIC apparatus hence leads to different types of formulae in different setups; see Claeskens & Hjort (2008, Chs. 5 and 6) for a fuller discussion, illustrations, and generalisations.

Hermansen & Hjort (2015) introduce a FIC for selection among nested parametric models for some classes of time series processes. The aim of this paper is to motivate an extension of this approach which will justify comparison and selection among both parametric and nonparametric candidate models. The derivation follows reasoning similar to the development of Jullum & Hjort (2015), where focused inference and model selection among parametric and nonparametric models are discussed for independent observations. By including a nonparametric candidate among the parametric models, we will in particular be able to detect whether our parametric models are off-target. The FIC can therefore be seen as an insurance mechanism against poorly specified parametric candidates. On the other hand, we usually achieve higher precision using parametric models when these are adequate.

The class of models we consider here are for zero-mean stationary Gaussian time series processes, say $\{Y_t\}$. The dependency structure, which in such cases determines the entire model, is completely specified by the covariance function $C(k) = \mathrm{Cov}(Y_{t+k}, Y_t)$, defined for all lags $k = 0, 1, 2, \ldots$. By Wold's theorem, see e.g. Priestley (1981), a function $C(k)$ is a proper covariance function if and only if there is a distribution G on $[-\pi, \pi]$, symmetric around zero, such that

$$C(k) = \int_{-\pi}^{\pi} e^{i\omega k}\,\mathrm{d}G(\omega) = 2\int_0^{\pi} \cos(\omega k)\,\mathrm{d}G(\omega) \quad \text{for } k = 0, 1, 2, \ldots. \tag{1}$$

We shall also take this spectral measure G to have a continuous density g, so that $C(k) = 2\int_0^{\pi} \cos(\omega k) g(\omega)\,\mathrm{d}\omega$. The spectral density can be obtained as the Fourier transform of the covariance function, see e.g. Brillinger (1975) and Dzhaparidze (1986). It is mathematically convenient to work in the frequency domain. Where necessary we write C_g to indicate that this is the covariance index by g obtained by (1).

The time series processes considered in Hermansen & Hjort (2015) were assumed to lie between well-defined narrow and wide models, in a local large-sample framework where the true spectral density f_{true} can be represented as $f_{\theta_0, \gamma_0 + \delta/\sqrt{n}}$; this framework causes variances and squared biases to be of the same order of magnitude $O(1/n)$. In the present paper we sidestep such conditions, however, going instead for direct assessment and then estimation of variances, squared biases, and hence mean squared errors.

We start out considering focus functions of the type

$$\mu(G, h) = \int_{-\pi}^{\pi} h(\omega)\,\mathrm{d}G(\omega) = 2\int_0^{\pi} h(\omega) g(\omega)\,\mathrm{d}\omega, \tag{2}$$

where h is a continuous and bounded function on $[-\pi, \pi]$, with potentially a finite number of jump discontinuities. This quite general class includes e.g. the covariance function. This construction allows also studying specific parts of the spectral density by using indicator functions; see also Gray (2006) for further illustration of quantities of type (2). Later results, employing also the delta method, will then make it easy to reach FIC formulae also for smooth functions of such $\mu(G, h)$, as correlations, threshold probabilities, etc.

A simple 'proof of concept' for focused model selection is presented in Figure 1. It pertains to a situation where the underlying (true) model is an autoregressive time series of order two with parameters $\sigma = 1.0$, $\rho_1 = 0.7$, $\rho_2 = -0.6$, and of length $n = 100$. The focus estimands considered are covariances $C(k)$ at lags $0, 1, 2, 3, 4, 5$. Five candidate models are examined: That of independence (autoregressive of order zero); the autoregressive of orders one and two; the moving average of order one; and finally the nonparametric one, where nothing more is assumed than saying that the series is stationary with a finite variance. The figure illustrates the crucial point that one and the same model is not necessarily best for all estimation purposes, and that there is a potential gain by including also the nonparametric alternative alongside parametric ones. See the figure text for details.

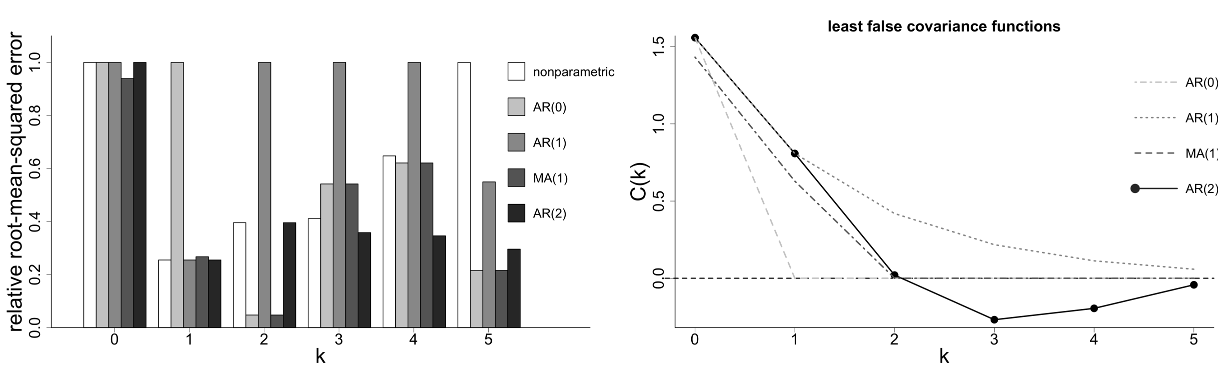

Figure 1: For the situation described above, simulation (with 5000 repetitions) is used to compute root-mse for the five candidate models, for each of the focus parameters $C(k)$. For ease of comparison we have scaled the root-mse to the unit interval. Note that since we have included the true model among our candidates, nonparametric estimation is never the optimal choice; it is however often close and it is the second best choice for lags 1 and 3. For lags 2 and 5, where the true values are close to zero, the simpler models, like AR(0) and MA(1), are highly successful, by achieving both low bias and low variance.

2. Estimation and approximations

We start out by investigating the behaviour of the two most common parametric estimation procedures, that based on maximum likelihood and that based on the Whittle approximation in (5).

2.1. Maximum likelihood estimation outside the model. Let $\underline{y}_n^{\rm t} = (y_1, \ldots, y_n)$ be a sample of size n from a zero mean stationary Gaussian time series process with spectral distribution function G and corresponding density g. Furthermore, let f_θ, with $\theta \in \Theta \subset \mathbb{R}^p$ for finite p, index an arbitrary parametric candidate. The corresponding full log-likelihood is $\ell_n(f_\theta) = -\frac{1}{2}\{n\log(2\pi) + \log|\Sigma_n(f_\theta)| + \underline{y}_n^{\rm t}\Sigma_n(f_\theta)^{-1}\underline{y}_n\}$, where $\Sigma_n(f_\theta)$ is the covariance matrix with elements $C_{f_\theta}(|i-j|)$ for $i,j = 1,\ldots,n$. Since the class of parametric candidate models is not assumed to necessarily include the true g, the maximum likelihood estimator does not converge to a 'true' parameter value. Instead it converges to the so-called 'least false' parameter value, i.e. $\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta) \to_{P_g} \arg\min_\theta d(g, f_\theta) = \theta_0$, where

$$d(g, f_\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log\frac{g(\omega)}{f_\theta(\omega)} + 1 - \frac{g(\omega)}{f_\theta(\omega)}\right){\rm d}\omega = -\frac{1}{4\pi}\int_{-\pi}^{\pi}(\log g(\omega) + 1)\,{\rm d}\omega - R(G,\theta), \tag{3}$$

see Dahlhaus & Wefelmeyer (1996). Moreover,

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d J_0^{-1}U, \qquad \text{where } U \sim {\rm N}_p(0, K_0), \tag{4}$$

with J_0 and K_0 defined by

$$J_0 = J(g, f_{\theta_0}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\Big[\nabla\Psi_{\theta_0}(\omega)\nabla\Psi_{\theta_0}(\omega)^{\rm t}g(\omega) + \nabla^2\Psi_{\theta_0}(\omega)\{f_{\theta_0}(\omega) - g(\omega)\}\Big]\frac{1}{f_{\theta_0}(\omega)}\,{\rm d}\omega$$

and

$$K_0 = K(g, f_{\theta_0}) = \frac{1}{4\pi}\int_{-\pi}^{\pi}\nabla\Psi_{\theta_0}(\omega)\nabla\Psi_{\theta_0}(\omega)^{\rm t}\left[\frac{g(\omega)}{f_{\theta_0}(\omega)}\right]^2{\rm d}\omega,$$

where $\Psi_\theta(\omega) = \log f_\theta(\omega)$ and $\nabla\Psi_\theta(\omega)$ and $\nabla^2\Psi_\theta(\omega)$ are the vector and matrix of partial derivatives with respect to θ, see Dahlhaus & Wefelmeyer (1996, Theorem 3.3).

2.2. The Whittle approximation. The Whittle pseudo log-likelihood is an approximation to the full Gaussian log-likelihood ℓ_n. It was originally suggested by P. Whittle in a series of works from the 1950s (cf. Whittle (1953)), and is defined as

$$\widetilde{\ell}_n(f) = -\frac{n}{2}\left\{\log 2\pi + \frac{1}{2\pi}\int_{-\pi}^{\pi}\log[2\pi f(\omega)]\,{\rm d}\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi}\frac{I_n(\omega)}{f(\omega)}\,{\rm d}\omega\right\}, \tag{5}$$

where $I_n(\omega) = (2\pi n)^{-1}|\sum_{t\le n} y_t\exp(i\omega t)|^2$ is the periodogram. This approximation is close to the full Gaussian log-likelihood in the sense that $\ell_n(f) = \widetilde{\ell}_n(f) + O_{P_g}(1)$ uniformly in f, see Coursol & Dacunha-Castelle (1982) or Dzhaparidze (1986) for details and additional discussion. More important here, however, is that (5) motivates an additional estimation procedure, namely the Whittle estimator $\widetilde{\theta}_n = \arg\min_\theta \widetilde{\ell}_n(f_\theta)$. This alternative estimator is easier to work with in practice (both analytically and numerically) and shares several properties with maximum likelihood estimator, e.g. $\sqrt{n}(\widetilde{\theta}_n - \theta_0)$ achieves the same limit distribution as in (4), with the same least false parameter value θ_0 as defined in relation to (3); see Dahlhaus & Wefelmeyer (1996) for details. This means that in a large-sample perspective, the maximum likelihood estimator and the simpler Whittle estimator are equally efficient and essentially interchangeable.

3. Parametric versus nonparametric

We shall now obtain large-sample approximations which can be used to construct estimates for the mse of the model based estimators for the focus parameter μ. This will in turn lead to FIC formulae.

3.1. How to compare parametric and nonparametric models? Let $\mu = \mu(G)$ be a focus function, i.e. a functional mapping of the spectral measure G to a scalar value. Often the collection of parametric candidate models, which we represent by F_θ, does not include the true G. The question is then which model should we use – parametric or nonparametric – for estimating the focus μ.

Let $\widehat{\mu}_{\rm np} = \mu(\widehat{G}_n)$ be the nonparametric estimate for the true $\mu_{\rm true} = \mu(G)$ and assume that

$$\sqrt{n}(\widehat{\mu}_{\rm np} - \mu_{\rm true}) \to_d {\rm N}(0, v_{\rm np}) \quad \text{and} \quad \sqrt{n}(\widehat{\mu}_{\rm pm} - \mu_0) \to_d {\rm N}(0, v_{\rm pm})$$

for each parametric candidate $\widehat{\mu}_{\rm pm} = \mu(F_{\widehat{\theta}_n})$, where $\mu_0 = \mu(F_{\theta_0})$ is the focus function evaluated under the least false model F_{θ_0} as discussed in relation to (3). Then, without going into details, the large-sample results above motivate the following first-order approximations for the mse of the estimated focus parameters:

$$\text{mse}_{\rm np} = 0^2 + v_{\rm np}/n = v_{\rm np}/n \quad \text{and} \quad \text{mse}_{\rm pm} = b^2 + v_{\rm pm}/n, \tag{6}$$

where $b = \mu_0 - \mu_{\rm true}$. The remainder of the section will be used to motivate and obtain good estimators for the mean squared errors in (6) with the class of foci $\mu(G, h)$ defined in (2).

3.2. Deriving unbiased risk estimates. In the derivation below, the parametric candidates F_θ will be fitted using the Whittle estimator $\widetilde{\theta}_n$ as defined in (5) and we will also use the periodogram based

$$\widetilde{G}_n = 2\int_0^\omega I_n(u)\,{\rm d}u \tag{7}$$

as a canonical estimator for the spectral measure G; see among others Taniguchi (1980). Using the Whittle estimator in collaboration with (7) results in a convenient simplification of the derivations below; extending the arguments to full ML estimation should be a straightforward extension by the techniques developed in Dahlhaus & Wefelmeyer (1996).

This motivates the following nonparametric and parametric estimators for $\mu(G, h)$ from (2):

$$\widetilde{\mu}_{\rm np} = \int_{-\pi}^{\pi} h(\omega) I_n(\omega)\,{\rm d}\omega = \frac{1}{n}\underline{y}_n^{\rm t}\Sigma_n(h)\underline{y}_n = X_n \quad \text{and} \quad \widetilde{\mu}_{\rm pm} = \int_{-\pi}^{\pi} h(\omega) f_{\widetilde{\theta}_n}(\omega)\,{\rm d}\omega.$$

The following proposition establishes the joint limit distribution for the estimators above (suitably normalised), which in turn will be used to obtain good approximations for their respective mean squared errors.

Proposition 1. *Let $y_1, \ldots, y_n$ be realisations from a stationary Gaussian time series model with spectral density g assumed to be uniformly bounded away from both zero and infinity. Suppose f_θ is two times differentiable with respect to θ and that f_θ, ∇f_θ and $\nabla^2 f_\theta$ are continuous and uniformly bounded in both ω and θ in a neighbourhood of the least false parameter value θ_0 as defined above (3). Then*

$$\begin{pmatrix} \sqrt{n}(\widetilde{\mu}_{\rm np} - \mu_{\rm true}) \\ \sqrt{n}(\widetilde{\mu}_{\rm pm} - \mu_0) \end{pmatrix} \to_d \begin{pmatrix} X \\ c^{\rm t} J(g, f_{\theta_0})^{-1} U \end{pmatrix} \sim {\rm N}_2\left(0, \begin{pmatrix} v_{\rm np} & v_c \\ v_c & v_{\rm pm} \end{pmatrix}\right), \tag{8}$$

where

$$v_{\rm np} = 4\pi \int_{-\pi}^{\pi} \{h(\omega) g(\omega)\}^2\,{\rm d}\omega \quad \text{and} \quad v_{\rm pm} = c^{\rm t} J(g, f_{\theta_0})^{-1} K(g, f_{\theta_0}) J(g, f_{\theta_0})^{-1} c,$$

with J and K as defined below (4), and $v_c = c^{\rm t} J(g, f_{\theta_0})^{-1} d$, where $c = \nabla\mu(f_{\theta_0})$ and

$$d = {\rm Cov}(X, U) = \int_{-\pi}^{\pi} \frac{\nabla f_{\theta_0}(\omega) h(\omega) g(\omega)^2}{f_{\theta_0}(\omega)^2}\,{\rm d}\omega.$$

Proof. It follows from the results in Dzhaparidze (1986) that $\widetilde{\theta}_n - \theta_0 = J(g, \theta_0)^{-1} U_n + o_{P_g}(1/\sqrt{n})$ where

$$U_n = \nabla\widetilde{\ell}_n(f_{\theta_0}) = -\frac{1}{2}\{{\rm tr}(\Sigma_n(\nabla\Psi_{\theta_0})) - \underline{y}_n^{\rm t}\Sigma_n(\nabla\Psi_{\theta_0}/f_\theta)\underline{y}_n\}$$

and $\Psi_{\theta_0} = \log f_{\theta_0}$ and $\nabla\Psi_{\theta_0}$ is the vector of partial derivatives. This means that the marginal distribution and the respective mean and variance are easily found by applying the standard delta method. Moreover, since $X_n = \underline{y}_n^{\rm t}\Sigma_n(h)\underline{y}_n/n$ and U_n are both quadratic forms, the joint limit distribution is readily obtainable by a Cramér–Wold type of argument; we will not go into details on this here, see Hermansen & Hjort (2014) for derivations of a similar type. To complete the proof, observe for the covariances that

$${\rm Cov}(X_n, U_n) = \frac{2}{n}{\rm tr}\{\Sigma_n(h)\Sigma_n(g)\Sigma_n(\nabla\Psi_\theta/f_\theta)\Sigma_n(g)\} \to \int_{-\pi}^{\pi} \frac{\nabla f_{\theta_0}(\omega) h(\omega) g(\omega)^2}{f(\omega)^2}\,{\rm d}\omega$$

from the results in Dzhaparidze (1986) or Dahlhaus & Wefelmeyer (1996, Lemma A.5). □

The nonparametric estimator is by construction unbiased in the limit; an estimate for the risk is therefore easily obtained from the variance formula above. For the parametric candidate, we need in addition an unbiased estimate for the squared bias. Following Jullum & Hjort (2015) we start with $\widetilde{b} = \widetilde{\mu}_{\rm pm} - \widetilde{\mu}_{\rm np}$ as an initial estimate for $b = \mu_0 - \mu_{\rm true}$. Since it follows from (8) that $\sqrt{n}(\widetilde{b} - b) \to_d c^t J^{-1} U - X \sim {\rm N}(0, \kappa)$, where $\kappa = v_{\rm pm} + v_{\rm np} - 2v_c$, we have ${\rm E}\,\widetilde{b}^2 \approx b^2 + \kappa/n + o(1/n)$. This leads to the following mse estimators

$$\mathrm{FIC}_{\rm np} = \widetilde{\rm mse}_{\rm np} = \widetilde{v}_{\rm np}/n \quad \text{and} \quad \mathrm{FIC}_{\rm pm} = \widetilde{\rm mse}_{\rm pm} = \widetilde{\rm bsq} + \widetilde{v}_{\rm pm}/n = \max(0, \widetilde{b}^2 - \widetilde{\kappa}/n) + \widetilde{v}_{\rm pm}/n.$$

With clear-cut estimates of the risk of the nonparametric and parametric models' estimators of μ, represented by the above FIC scores, our model selection strategy is as follows: Compute the FIC score for each of the say m parametric candidate models and for the nonparametric alternative; rank these $m+1$ scores; and select the model and estimator associated with the smallest FIC score. Note that the same $\mathrm{FIC}_{\rm pm}$ formula (with different estimates and quantities) is used for all the different parametric candidate models.
In Figure 2 we revisit the lagged covariance estimation problem of the introduction, to see how the FIC performs compared to the AIC, BIC and always using the nonparametric estimate. Note that the AIC and BIC tools do not work for the nonparametric model, since there is no likelihood function.

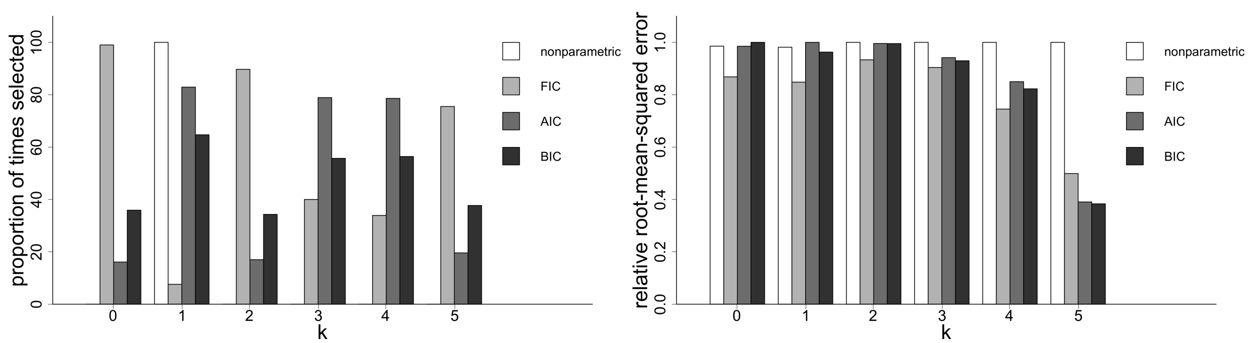

Figure 2: This is a continuation of the illustration in Figure 1. In the right panel we have calculated (using simulations) the number of times each criterion, i.e. always nonparametric, FIC, AIC and BIC selects the model that has the theoretical lowest root-mean-squared error. Note that AIC and BIC only selects among the parametric models. For lag 1 the theoretical root-mse for the autoregressive models are, for all practical purposes, equal to what obtained by the nonparametric model and all are therefore seen as attaining the smallest root-mse. In the right panel we have compared (in the same simulations) the obtained root-mse for the same four scenarios. Here we see that the FIC behaves as intended by selecting (on average) the models that produces the smallest risk.

Although we have concentrated on focus functions $\mu(G, h)$ given by (2), our focused model selection strategy is not restricted to this class, and should work for any focus functions fulfilling (8). Hence, a standard delta method argument ensures that any continuously differentiable function of a finite set of focus parameters $\mu_1, \ldots, \mu_q$ all on the form of (2), may be handled by our scheme – provided there is joint convergence for all individual estimators. For instance, this allows for focus functions like the lag k correlation function $\mu = \rho(k) = C(k)/C(0)$. A further extension of this is to replace (2) by $\mu^\circ(g, H) = \int_{-\pi}^{\pi} H(g(\omega))\,{\rm d}\omega$, where H is continuous on the (finite) range of g; see among others Taniguchi (1980). In completely general terms, our results may be generalised to focus functions $\mu = T(G)$ for well behaved functionals T mapping the spectral distribution G to a scalar value. It is also possible to extend the arguments to other parametric estimation procedures, especially if they are derived as minimisers of the empirical analogue of $\arg\min_\theta R(G, \theta)$ for R the model specific part of possibly different divergence measure than in (3), see Dahlhaus & Wefelmeyer (1996) and Taniguchi (1981) for alternatives.

4. Concluding remarks

A. AFIC. Our discussion has so far been in terms of a given focus parameter μ. In situations where several focus parameters are considered jointly, say correlations of orders 1 to 5, the FIC machinery can easily be lifted to one involving a weighted average of FIC scores, the AFIC, with weights reflecting importance dictated by the statistician.

B. Model averaging. The FIC scores may also be used to combine the most promising estimators into a model averaged estimator, say $\widehat{\mu}^* = \sum_{\rm models} c(M_j)\widehat{\mu}_j$, with $c(M_j)$ given higher values for models with good FIC scores.

C. The conditional FIC. For time series processes several interesting and important foci are naturally related to predictions, sample size dependent or formulated conditional on past observations, e.g. h-step ahead predictions but also $\mu(y_1, \ldots, y_m) = \Pr\{Y_{n+1} > \alpha \text{ and } Y_{n+2} > \alpha \,|\, y_1, \ldots, y_m\}$ for a suitable choice of α. The dependency on previous data requires a new and extended modelling framework, which in Hermansen & Hjort (2015, Sections 5 & 6) led to generalisations and also motivated the conditional focused information criterion (cFIC). These considerations need to be taken properly into account in a complete extension of the FIC methodology for time series in the framework of selecting among parametric and nonparametric models.

D. Trends and covariates. Methods and results of our paper may be generalised to classes of models of the type $Y_t = m(t, \beta) + x_i^{\rm t}\gamma + \varepsilon_t$, with ε_t a stationary Gaussian time process. These issues, leading to a larger repertoire of FIC formulae, will be returned to in later work.

References

Brillinger, D. (1975). *Time Series: Data Analysis and Theory.* Holt, Rinehart and Winston.

Claeskens, G. & Hjort, N. (2003). The focused information criterion [with discussion and rejoinder]. *Journal of the American Statistical Association*, 98, 900–916.

Claeskens, G. & Hjort, N. (2008). *Model Selection and Model Averaging.* Cambridge University Press.

Coursol, J. & Dacunha-Castelle, D. (1982). Remarks on the approximation of the likelihood function of a stationary Gaussian process. *Theory of Probab. Appl.*, 27, 162–167.

Dahlhaus, R. & Wefelmeyer, W. (1996). Asymptotically optimal estimation in misspecified time series models. *Annals of Statistics*, 24, 952–973.

Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series.* Springer.

Gray, R. (2006). *Toeplitz and Circulant Matrices: A Review.* Now publishers Inc.

Hermansen, G. & Hjort, N. (2014). *Limiting normality of quadratic forms, with applications to time series analysis.* Technical report, University of Oslo and Norwegian Computing Centre.

Hermansen, G. & Hjort, N. (2015). Focused information criteria for time series. *Submitted for publication.*

Jullum, M. & Hjort, N. (2015). Parametric or nonparametric: The FIC approach. *Submitted for publication.*

Priestley, M. (1981). *Spectral Analysis and Time Series.* Academic Press.

Taniguchi, M. (1980). On estimation of the integrals of certain functions of spectral density. *Journal of Applied Probability*, 17, 73–80.

Taniguchi, M. (1981). An estimation procedure of parameters of a certain spectral density model. *Journal of the Royal Statistical Society. Series B (Methodological)*, 43, 34–40.

Whittle, P. (1953). The analysis of multiple stationary time series. *Journal of the Royal Statistical Society. Series B (Methodological)*, 15, 125–139.