Parametric or Nonparametric: The FIC Approach for time series

Gudmund Horn Hermansen

Nils Lid Hjort

Martin Jullum

University of Oslo

martinju@math.uio.no

June 17, 2015

Modelling time series

Working class

- Zero-mean stationary Gaussian time series, $Y_t, t = 1, \ldots, n$
- The model is fully specified by the covariance function:

$$C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}) = \int_{-\pi}^{\pi} \exp(i\omega k)\, \mathrm{d}G(\omega) = 2\int_0^{\pi} \cos(\omega k)\, \mathrm{d}G(\omega)$$

- Spectral measure G with spectral density g unknown, how should we model it?
- Parametric approach: Use a spectral measure F_θ with spectral density f_θ, parameterized by some θ
  - AR(p), MA(q), ARMA(p, q), etc.
- Nonparametric approach: Periodogram based $\widetilde{G}_n = 2\int_0^{\omega} I_n(u)\, \mathrm{d}u$, for $I_n(\omega) = \frac{1}{2\pi n}\left|\sum_{t \leq n} y_t \exp(i\omega t)\right|^2$ the periodogram

Model selection for time series

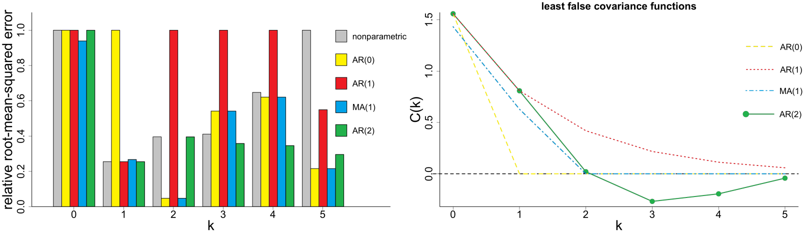
- Classical model selection
  - Information criteria: AIC, BIC, AICc, FPE , HQ-criterion, ...
  - Goodness-of-fit/ badness-of-fit test: Testing 'whiteness' of fitted residuals with Portmanteau tests, frequency domain tests, etc.
- No particularly well-suited criterion for model selection among parametrics and nonparametrics
- Different models have strengths and weaknesses at different parts of the data space
- What you want to learn should reflect your choice of model

Model selection for time series

- Classical model selection
  - Information criteria: AIC, BIC, AICc, FPE , HQ-criterion, ...
  - Goodness-of-fit/ badness-of-fit test: Testing 'whiteness' of fitted residuals with Portmanteau tests, frequency domain tests, etc.
- No particularly well-suited criterion for model selection among parametrics and nonparametrics
- Different models have strengths and weaknesses at different parts of the data space
- What you want to learn should reflect your choice of model

Proof of concept – estimate $C_G(k)$

- Simulate AR(2) with $n = 100$ where $\sigma = 1, \rho = (0.7, -0.6)^t$
- Performance for 5 different candidate models: AR(0), AR(1), AR(2), MA(1), Nonparametric



- Left panel: Root-MSE of model based estimators for different lags (relative to worst estimate)
- Right panel: Limiting covariance functions for different models
- Different models are best for different lags

Focus parameter

- A general population quantity μ of interest

Examples

- Covariance/correlation lag k:
  $C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}), \quad \rho_G(k) = \mathrm{corr}(Y_t, Y_{t+k})$
- Conditional expectation k step ahead $\mathrm{E}_G[Y_{t+k}|Y_t, \ldots, Y_{t-m}]$
- Threshold probabilities: $\Pr_G\{Y_{t+k} > y_0|Y_t, \ldots, Y_{t-m}\}$

- A functional of the spectral measure G, for simplicity assumed to be on the form

$$\mu(G, s, h) = s\left(\int_{-\pi}^{\pi} h(\omega)\,\mathrm{d}G(\omega)\right),$$

where $h^{\mathrm{t}} = (h_1, \ldots, h_m)$, with the h_j are univariate functions bounded on $[-\pi, \pi]$ with a finite number of discontinuities and s is smooth function from $\mathbb{R}^m$ to $\mathbb{R}$

Focus parameter

- A general population quantity μ of interest

Examples

- Covariance/correlation lag k:
  $C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}), \quad \rho_G(k) = \mathrm{corr}(Y_t, Y_{t+k})$
- Conditional expectation k step ahead $\mathrm{E}_G[Y_{t+k}|Y_t, \ldots, Y_{t-m}]$
- Threshold probabilities: $\Pr_G\{Y_{t+k} > y_0|Y_t, \ldots, Y_{t-m}\}$
- A functional of the spectral measure G, for simplicity assumed to be on the form

$$\mu(G, s, h) = s\left(\int_{-\pi}^{\pi} h(\omega)\,\mathrm{d}G(\omega)\right),$$

  where $h^{\mathrm{t}} = (h_1, \ldots, h_m)$, with the h_j are univariate functions bounded on $[-\pi, \pi]$ with a finite number of discontinuities and s is smooth function from $\mathbb{R}^m$ to $\mathbb{R}$

Focus parameter

- A general population quantity μ of interest

Examples

- Covariance/correlation lag k:
  $C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}), \quad \rho_G(k) = \mathrm{corr}(Y_t, Y_{t+k})$
- Conditional expectation k step ahead $\mathrm{E}_G[Y_{t+k}|Y_t, \ldots, Y_{t-m}]$
- Threshold probabilities: $\Pr_G\{Y_{t+k} > y_0|Y_t, \ldots, Y_{t-m}\}$
- A functional of the spectral measure G, for simplicity assumed to be on the form

$$\mu(G, s, h) = s\left(\int_{-\pi}^{\pi} h(\omega)\,\mathrm{d}G(\omega)\right),$$

where $h^{\mathrm{t}} = (h_1, \ldots, h_m)$, with the h_j are univariate functions bounded on $[-\pi, \pi]$ with a finite number of discontinuities and s is smooth function from $\mathbb{R}^m$ to $\mathbb{R}$

Focus parameter

- A general population quantity μ of interest

Examples

- Covariance/correlation lag k:
  $C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}), \quad \rho_G(k) = \mathrm{corr}(Y_t, Y_{t+k})$
- Conditional expectation k step ahead $\mathrm{E}_G[Y_{t+k}|Y_t, \ldots, Y_{t-m}]$
- Threshold probabilities: $\mathrm{Pr}_G\{Y_{t+k} > y_0|Y_t, \ldots, Y_{t-m}\}$

- A functional of the spectral measure G, for simplicity assumed to be on the form

$$\mu(G, s, h) = s\left(\int_{-\pi}^{\pi} h(\omega)\, \mathrm{d}G(\omega)\right),$$

where $h^{\mathrm{t}} = (h_1, \ldots, h_m)$, with the h_j are univariate functions bounded on $[-\pi, \pi]$ with a finite number of discontinuities and s is smooth function from $\mathbb{R}^m$ to $\mathbb{R}$

Focus parameter

- A general population quantity μ of interest

Examples

- Covariance/correlation lag k:
  $C_G(k) = \mathrm{Cov}(Y_t, Y_{t+k}), \quad \rho_G(k) = \mathrm{corr}(Y_t, Y_{t+k})$
- Conditional expectation k step ahead $\mathrm{E}_G[Y_{t+k}|Y_t, \dots, Y_{t-m}]$
- Threshold probabilities: $\Pr_G\{Y_{t+k} > y_0 | Y_t, \dots, Y_{t-m}\}$

- A functional of the spectral measure G, for simplicity assumed to be on the form

$$\mu(G, s, h) = s\left(\int_{-\pi}^{\pi} h(\omega)\, \mathrm{d}G(\omega)\right),$$

where $h^{\mathrm{t}} = (h_1, \dots, h_m)$, with the h_j are univariate functions bounded on $[-\pi, \pi]$ with a finite number of discontinuities and s is smooth function from $\mathbb{R}^m$ to $\mathbb{R}$

Criterion idea

- Model selection problem $\rightarrow$ Best estimator for μ
- Estimate μ by plug-in estimation for each model M: $\widehat{\mu}_M = \mu(\widehat{G}_M, s, h)$
- Performance measure:

$$\mathrm{mse}(\widehat{\mu}_M) = \mathsf{E}\left[(\widehat{\mu}_M - \mu_{\mathrm{true}})^2\right] = \mathrm{bias}^2(\widehat{\mu}_M) + \mathrm{Var}(\widehat{\mu}_M)$$

Basic idea

- Estimate the mean squared error (mse) as squared bias + variance:

$$\mathrm{FIC}(M) = \widehat{\mathrm{mse}}(\widehat{\mu}_M) = \widehat{\mathrm{bias}^2}(\widehat{\mu}_M) + \widehat{\mathrm{Var}}(\widehat{\mu}_M)$$

- Choose the model and estimator with the smallest estimated mse

Criterion idea

- Model selection problem $\rightarrow$ Best estimator for μ
- Estimate μ by plug-in estimation for each model M: $\widehat{\mu}_M = \mu(\widehat{G}_M, s, h)$
- Performance measure:

$$\mathrm{mse}(\widehat{\mu}_M) = \mathsf{E}\left[(\widehat{\mu}_M - \mu_{\mathrm{true}})^2\right] = \mathrm{bias}^2(\widehat{\mu}_M) + \mathrm{Var}(\widehat{\mu}_M)$$

Basic idea

- Estimate the mean squared error (mse) as squared bias + variance:

$$\mathsf{FIC}(M) = \widehat{\mathrm{mse}}(\widehat{\mu}_M) = \widehat{\mathrm{bias}^2}(\widehat{\mu}_M) + \widehat{\mathrm{Var}}(\widehat{\mu}_M)$$

- Choose the model and estimator with the smallest estimated mse

Estimation and limit theory

Parametric estimation (Maximum likelihood):

- ML-estimator: $\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta)$, for $\ell_n(\theta)$ Gaussian log-likelihood of parametric spectral measure class F_θ
- No θ_{true} exists outside model conditions. Rather θ_0 least false parameter minimizing $d(g, f_\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log\frac{g(\omega)}{f_\theta(\omega)} + 1 - \frac{g(\omega)}{f_\theta(\omega)}\right)\mathrm{d}\omega$
- Asymptotic distribution: $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, \Sigma_0)$, with $\Sigma_0 = \Sigma_0(g, f_{\theta_0})$

$$\Rightarrow \sqrt{n}(F_{\widehat{\theta}_n}(\omega) - F_{\theta_0}(\omega)) \to_d X(\omega),$$

for $\omega \in [0, \pi]$ and $X(\cdot)$ a zero-mean Gaussian process

Nonparametric estimation

- Estimator of spectral measure G: $\widetilde{G}_n(\omega) = 2\int_0^\omega I_n(\omega)\,\mathrm{d}\omega$
- Asymptotic distribution:

$$\sqrt{n}(\widetilde{G}_n(\omega) - G(\omega)) \to_d W_2\left(2\pi\int_0^\omega g(u)^2\,\mathrm{d}u\right),$$

for $\omega \in [0, \pi]$, where $W(\cdot)$ is a standard Wiener process

Estimation and limit theory

Parametric estimation (Maximum likelihood):

- ML-estimator: $\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta)$, for $\ell_n(\theta)$ Gaussian log-likelihood of parametric spectral measure class F_θ
- No θ_{true} exists outside model conditions. Rather θ_0 least false parameter minimizing $d(g, f_\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log\frac{g(\omega)}{f_\theta(\omega)} + 1 - \frac{g(\omega)}{f_\theta(\omega)}\right)\mathrm{d}\omega$
- Asymptotic distribution: $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, \Sigma_0)$, with $\Sigma_0 = \Sigma_0(g, f_{\theta_0})$

$$\Rightarrow \sqrt{n}(F_{\widehat{\theta}_n}(\omega) - F_{\theta_0}(\omega)) \to_d X(\omega),$$

for $\omega \in [0, \pi]$ and $X(\cdot)$ a zero-mean Gaussian process

Nonparametric estimation

- Estimator of spectral measure G: $\widetilde{G}_n(\omega) = 2\int_0^\omega I_n(\omega)\,\mathrm{d}\omega$
- Asymptotic distribution:

$$\sqrt{n}(\widetilde{G}_n(\omega) - G(\omega)) \to_d W_2\left(2\pi\int_0^\omega g(u)^2\,\mathrm{d}u\right),$$

for $\omega \in [0, \pi]$, where $W(\cdot)$ is a standard Wiener process

Estimation and limit theory

Parametric estimation (Maximum likelihood):

- ML-estimator: $\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta)$, for $\ell_n(\theta)$ Gaussian log-likelihood of parametric spectral measure class F_θ
- No θ_{true} exists outside model conditions. Rather θ_0 least false parameter minimizing $d(g, f_\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log\frac{g(\omega)}{f_\theta(\omega)} + 1 - \frac{g(\omega)}{f_\theta(\omega)}\right)\mathrm{d}\omega$
- Asymptotic distribution: $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, \Sigma_0)$, with $\Sigma_0 = \Sigma_0(g, f_{\theta_0})$

$$\Rightarrow \sqrt{n}(F_{\widehat{\theta}_n}(\omega) - F_{\theta_0}(\omega)) \to_d X(\omega),$$

for $\omega \in [0, \pi]$ and $X(\cdot)$ a zero-mean Gaussian process

Nonparametric estimation

- Estimator of spectral measure G: $\widetilde{G}_n(\omega) = 2\int_0^\omega I_n(\omega)\,\mathrm{d}\omega$
- Asymptotic distribution:

$$\sqrt{n}(\widetilde{G}_n(\omega) - G(\omega)) \to_d W_2\left(2\pi\int_0^\omega g(u)^2\,\mathrm{d}u\right),$$

for $\omega \in [0, \pi]$, where $W(\cdot)$ is a standard Wiener process

Estimation and limit theory

Parametric estimation (Maximum likelihood):

- ML-estimator: $\widehat{\theta}_n = \arg\max_\theta \ell_n(\theta)$, for $\ell_n(\theta)$ Gaussian log-likelihood of parametric spectral measure class F_θ
- No θ_{true} exists outside model conditions. Rather θ_0 least false parameter minimizing $d(g, f_\theta) = -\frac{1}{4\pi}\int_{-\pi}^{\pi}\left(\log\frac{g(\omega)}{f_\theta(\omega)} + 1 - \frac{g(\omega)}{f_\theta(\omega)}\right)\mathrm{d}\omega$
- Asymptotic distribution: $\sqrt{n}(\widehat{\theta}_n - \theta_0) \to_d \mathrm{N}(0, \Sigma_0)$, with $\Sigma_0 = \Sigma_0(g, f_{\theta_0})$

$$\Rightarrow \sqrt{n}(F_{\widehat{\theta}_n}(\omega) - F_{\theta_0}(\omega)) \to_d X(\omega),$$

for $\omega \in [0, \pi]$ and $X(\cdot)$ a zero-mean Gaussian process

Nonparametric estimation

- Estimator of spectral measure G: $\widetilde{G}_n(\omega) = 2\int_0^\omega I_n(\omega)\,\mathrm{d}\omega$
- Asymptotic distribution:

$$\sqrt{n}(\widetilde{G}_n(\omega) - G(\omega)) \to_d W_2\left(2\pi\int_0^\omega g(u)^2\,\mathrm{d}u\right),$$

for $\omega \in [0, \pi]$, where $W(\cdot)$ is a standard Wiener process

MSE approximation

- $\widehat{\mu}_{\mathrm{pm}} = \mu(F_{\widehat{\theta}_n}, s, h) = s(\int_{-\pi}^{\pi} h(\omega) f_{\widehat{\theta}_n}(\omega)\,\mathrm{d}\omega)$
- $\widehat{\mu}_{\mathrm{np}} = \mu(\widetilde{G}_n, s, h) = s(\int_{-\pi}^{\pi} h(\omega) I_n(\omega)\,\mathrm{d}\omega)$
- Continuous mapping theorem, delta method etc gives:

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_{\mathrm{pm}} - \mu_0 \\ \widehat{\mu}_{\mathrm{np}} - \mu_{\mathrm{true}} \end{pmatrix} \to_d \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{\mathrm{pm}} & v_c \\ v_c & v_{\mathrm{np}} \end{pmatrix}\right),$$

where $\mu_0 = \mu(F_{\theta_0}, s, h)$ and $\mu_{\mathrm{true}} = \mu(G, s, h)$

- First order approximations:

$$\mathrm{mse}_{\mathrm{pm}} \approx b^2 + v_{\mathrm{pm}}/n \quad \text{and} \quad \mathrm{mse}_{\mathrm{np}} \approx 0^2 + v_{\mathrm{np}}/n,$$

for $b = \mu_0 - \mu_{\mathrm{true}}$

MSE approximation

- $\widehat{\mu}_{\text{pm}} = \mu(F_{\widehat{\theta}_n}, s, h) = s(\int_{-\pi}^{\pi} h(\omega) f_{\widehat{\theta}_n}(\omega)\,\mathrm{d}\omega)$
- $\widehat{\mu}_{\text{np}} = \mu(\widetilde{G}_n, s, h) = s(\int_{-\pi}^{\pi} h(\omega) I_n(\omega)\,\mathrm{d}\omega)$
- Continuous mapping theorem, delta method etc gives:

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_{\text{pm}} - \mu_0 \\ \widehat{\mu}_{\text{np}} - \mu_{\text{true}} \end{pmatrix} \to_d \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{\text{pm}} & v_c \\ v_c & v_{\text{np}} \end{pmatrix}\right),$$

where $\mu_0 = \mu(F_{\theta_0}, s, h)$ and $\mu_{\text{true}} = \mu(G, s, h)$

- First order approximations:

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

for $b = \mu_0 - \mu_{\text{true}}$

MSE approximation

- $\widehat{\mu}_{\rm pm} = \mu(F_{\widehat{\theta}_n}, s, h) = s(\int_{-\pi}^{\pi} h(\omega) f_{\widehat{\theta}_n}(\omega)\,\mathrm{d}\omega)$
- $\widehat{\mu}_{\rm np} = \mu(\widetilde{G}_n, s, h) = s(\int_{-\pi}^{\pi} h(\omega) I_n(\omega)\,\mathrm{d}\omega)$
- Continuous mapping theorem, delta method etc gives:

$$\sqrt{n}\begin{pmatrix} \widehat{\mu}_{\rm pm} - \mu_0 \\ \widehat{\mu}_{\rm np} - \mu_{\rm true} \end{pmatrix} \to_d \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} v_{\rm pm} & v_c \\ v_c & v_{\rm np} \end{pmatrix}\right),$$

where $\mu_0 = \mu(F_{\theta_0}, s, h)$ and $\mu_{\rm true} = \mu(G, s, h)$

- First order approximations:

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

for $b = \mu_0 - \mu_{\rm true}$

MSE estimation and the FIC

Estimate unknown quantities by empirical analogues

$$v_{\mathrm{pm}} \approx \widehat{v}_{\mathrm{pm}}, \quad v_{\mathrm{np}} \approx \widehat{v}_{\mathrm{np}},$$

and

$$b^2 \approx \max\{0, \widehat{b}^2 - \widehat{v}_b/n\},$$

where

$$\widehat{b} = \widehat{\mu}_{\mathrm{pm}} - \widehat{\mu}_{\mathrm{np}} \quad \text{and} \quad \widehat{v}_b = \widehat{v}_{\mathrm{pm}} + \widehat{v}_{\mathrm{np}} - 2\widehat{v}_c$$

Mse estimates: FIC scores

$$\mathrm{FIC}(\widehat{\mu}_{\mathrm{pm}}) = \max\{0, \widehat{b}^2 - \widehat{v}_b\} + \widehat{v}_{\mathrm{pm}}/n$$
$$\mathrm{FIC}(\widehat{\mu}_{\mathrm{np}}) = \widehat{v}_{\mathrm{np}}/n$$

- Compute for all parametric candidates + nonparametric
- Choose the model and estimator with the smallest FIC-score

MSE estimation and the FIC

Estimate unknown quantities by empirical analogues

$$v_{\rm pm} \approx \widehat{v}_{\rm pm}, \quad v_{\rm np} \approx \widehat{v}_{\rm np},$$

and

$$b^2 \approx \max\{0, \widehat{b}^2 - \widehat{v}_b/n\},$$

where

$$\widehat{b} = \widehat{\mu}_{\rm pm} - \widehat{\mu}_{\rm np} \quad \text{and} \quad \widehat{v}_b = \widehat{v}_{\rm pm} + \widehat{v}_{\rm np} - 2\widehat{v}_c$$

Mse estimates: FIC scores

$$\mathrm{FIC}(\widehat{\mu}_{\rm pm}) = \max\{0, \widehat{b}^2 - \widehat{v}_b\} + \widehat{v}_{\rm pm}/n$$

$$\mathrm{FIC}(\widehat{\mu}_{\rm np}) = \widehat{v}_{\rm np}/n$$

- Compute for all parametric candidates + nonparametric
- Choose the model and estimator with the smallest FIC-score

Theoretic properties

- Estimation outside model conditions
- Consistent variance and squared bias estimators

FIC asymptotics

- Parametrics biased: $Pr\{\text{Select pm}\} \to 0$
- Parametrics correct: $Pr\{\text{Select pm}\} \to \chi_1^2(2) \approx 0.843$
- Implicit focused hypothesis test of parametric model with significance level 0.157

Theoretic properties

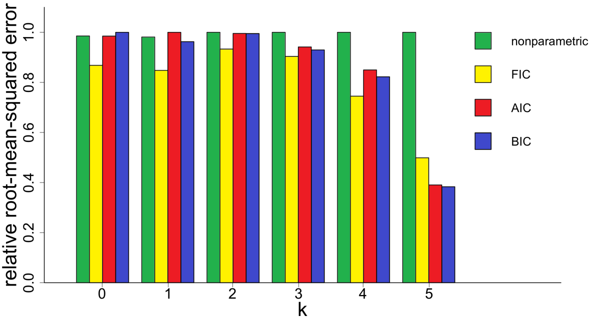
- Estimation outside model conditions
- Consistent variance and squared bias estimators

FIC asymptotics

- Parametrics biased: $Pr\,\{\text{Select pm}\} \to 0$
- Parametrics correct: $Pr\,\{\text{Select pm}\} \to \chi_1^2(2) \approx 0.843$
- Implicit focused hypothesis test of parametric model with significance level 0.157

Performance comparison

- Revisit the lagged covariance example
- Compare empirical root mean squared error of the estimators selected by FIC, AIC, BIC and Nonparametric



- The FIC works as intended choosing models with smaller risk for estimation of the different lags

Concluding remarks

- AFIC: Consider a weighted set of focus parameters simulatenously
  - Minimize risk $= \int \mathrm{mse}(\widehat{\mu}(t))\,\mathrm{d}W(t)$
  - Conceptually solved using $\int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathrm{d}W(t)$
- Time series with trends and covariates: $Y_t = m(t;\beta) + x_i^{\mathrm{t}}\gamma + \varepsilon_i$, with ε_i a zero-mean Gaussian stationary time series process

Origin:

- Claeskens and Hjort (2003): The Focussed Information Criterion
- Hermansen and Hjort (2014): Focused Information Criteria for time series
- Jullum and Hjort (2015): Parametric or Nonparametric: The FIC Approach

Concluding remarks

- AFIC: Consider a weighted set of focus parameters simulatenously
  - Minimize risk $= \int \mathrm{mse}(\widehat{\mu}(t))\,\mathrm{d}W(t)$
  - Conceptually solved using $\int \mathrm{FIC}(\widehat{\mu}_M(t))\,\mathrm{d}W(t)$
- Time series with trends and covariates: $Y_t = m(t;\beta) + x_i^{\mathrm{t}}\gamma + \varepsilon_i$, with ε_i a zero-mean Gaussian stationary time series process

Origin:

- Claeskens and Hjort (2003): The Focussed Information Criterion
- Hermansen and Hjort (2014): Focused Information Criteria for time series
- Jullum and Hjort (2015): Parametric or Nonparametric: The FIC Approach