Frequency Domain Inference for Seasonally Persistent Processes

David A. Stephens

d.stephens@imperial.ac.uk

Department of Mathematics

Imperial College London

5th January, 2006

Joint work with Sofia Olhede and Emma McCoy
Department of Mathematics, Imperial College London.

Outline

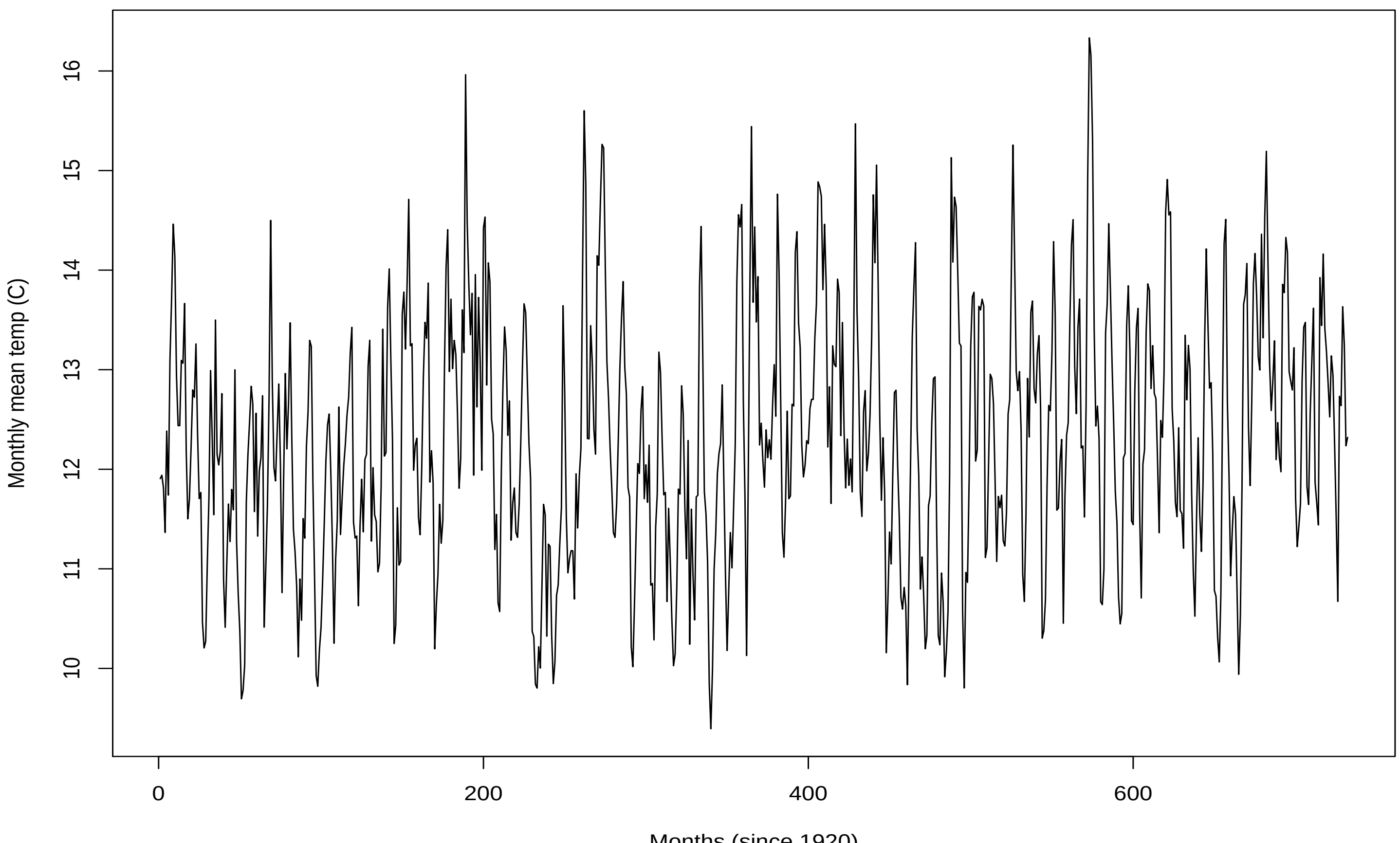
- Motivation and Notation
- ARMA models
- Frequency Domain Inference and The Whittle Likelihood
- Persistence
- Gegenbauer models
- A Whittle likelihood for Seasonal Processes
- Some results
- Some examples

Some real data examples

- Farallon data
- Carinae data
- SOI data
- CO_2 data
- Quebec Births Data

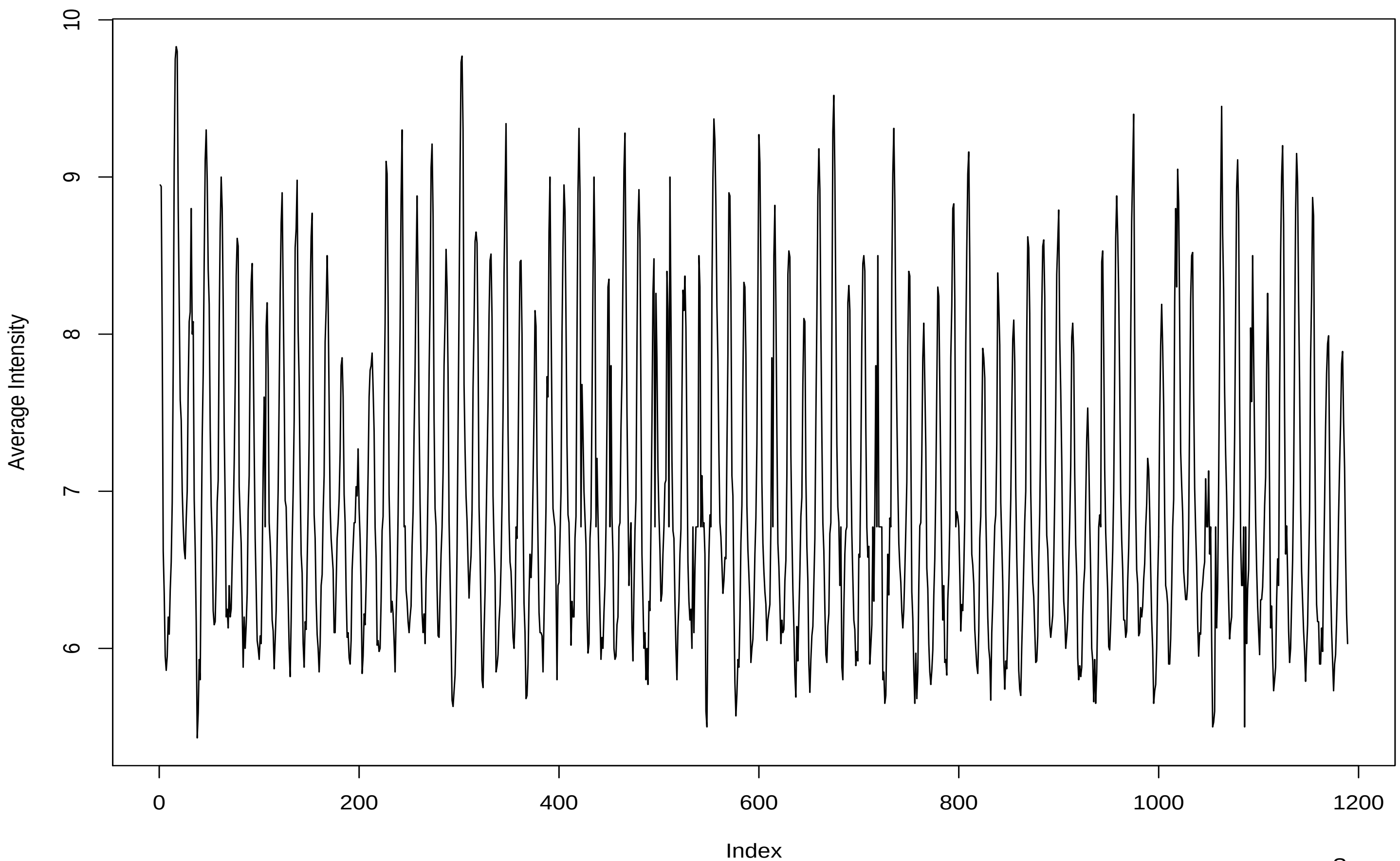
Example : Farallon data

732 obs. monthly mean temperature at Farallon Islands, Ca, 1920-1981



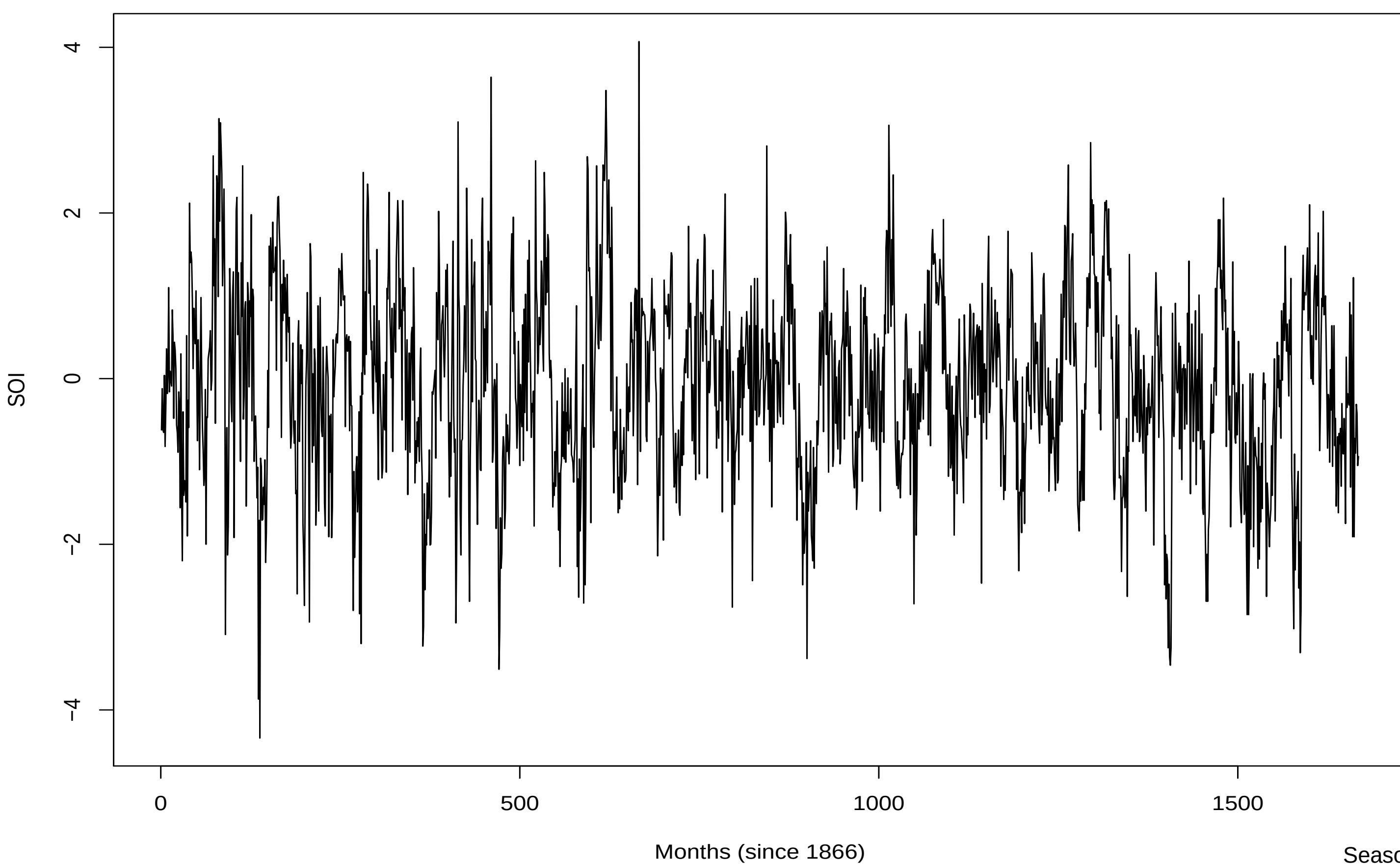
Example : Carinae data

1182 consecutive 10-day mean light intensities of the *S. Carinæ* variable star.



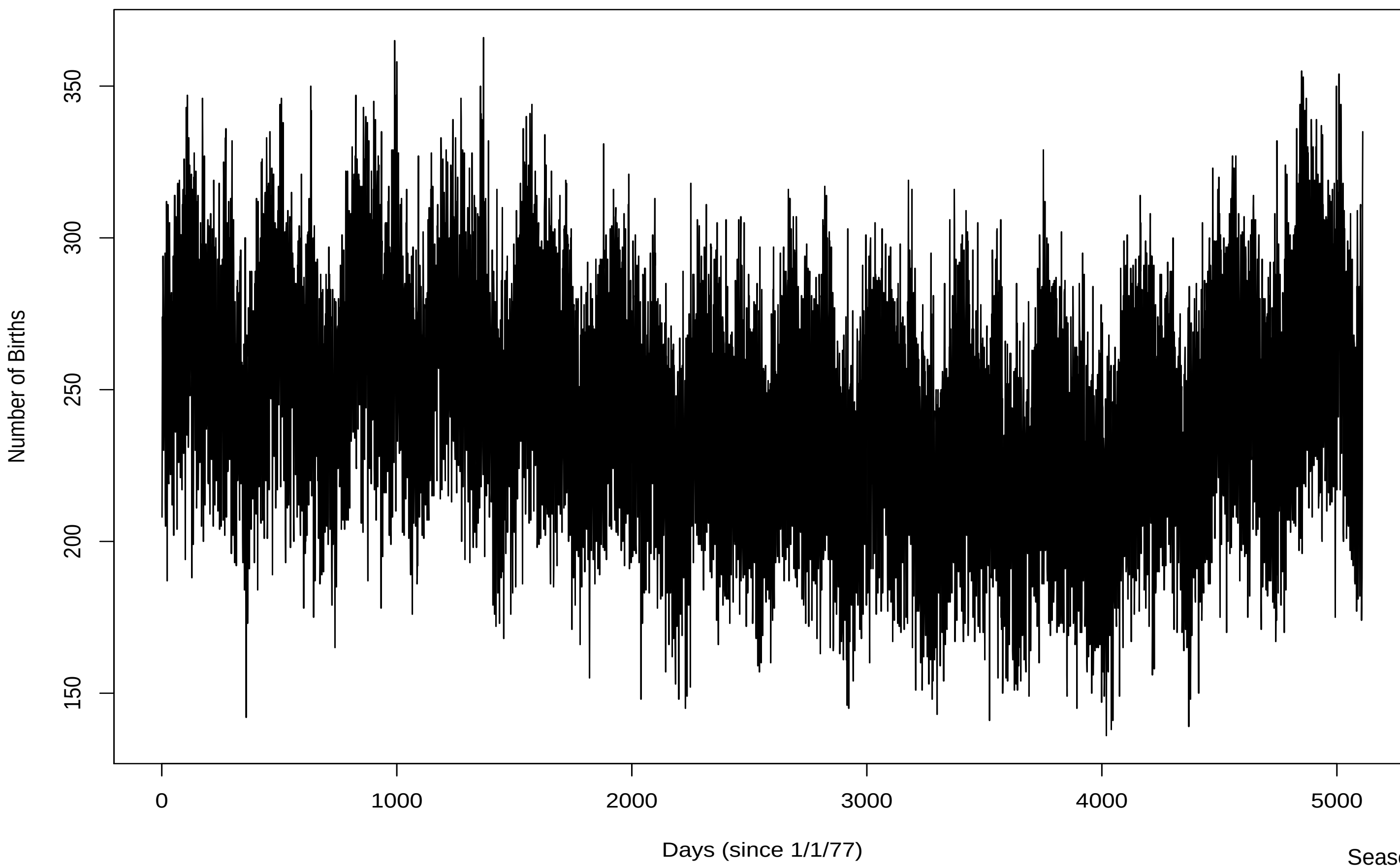
Example : SOI data

1688 monthly values of SOI (normalized pressure difference Tahiti/Darwin: `www.cru.uea.ac.uk`)



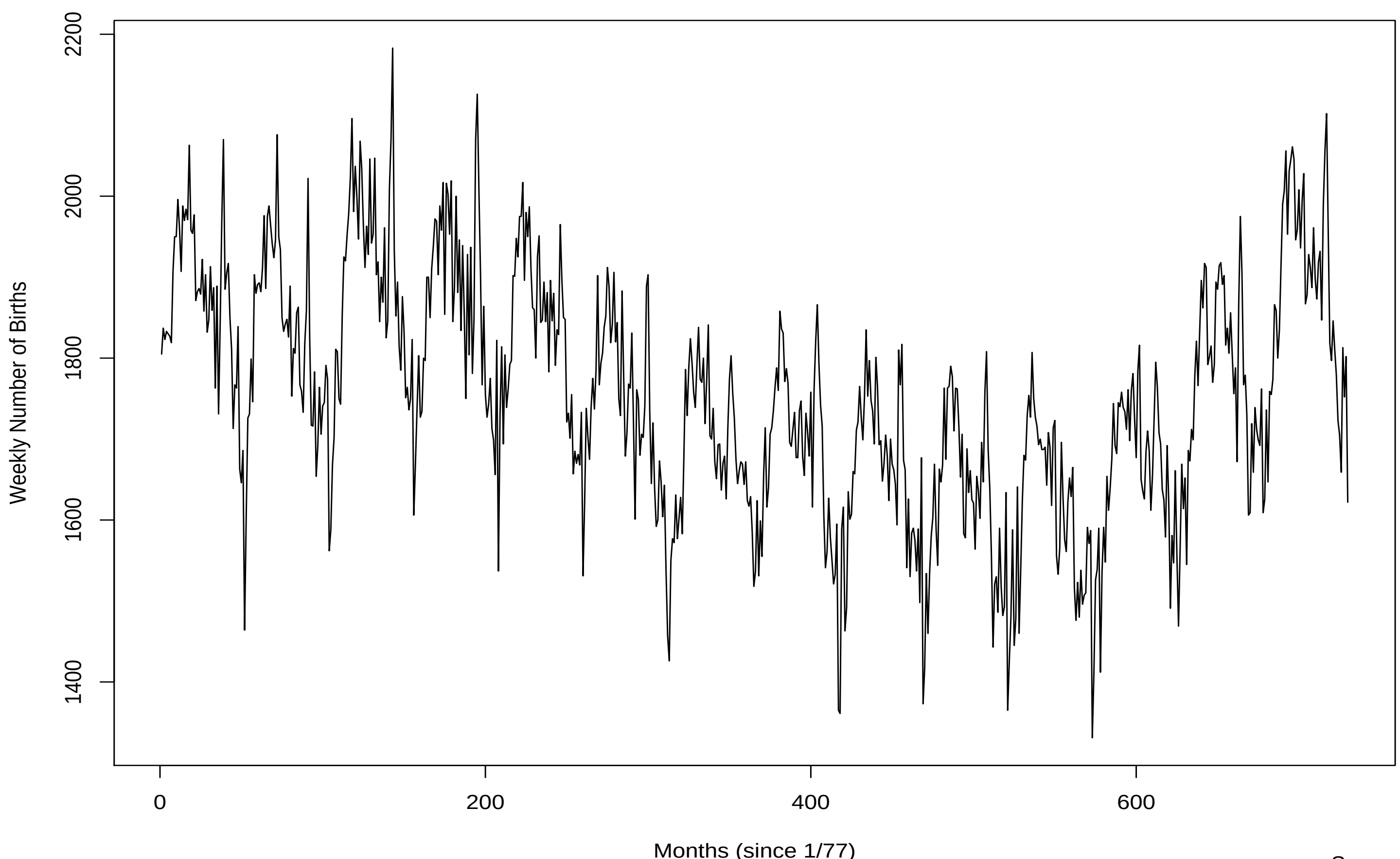
Example : Quebec Births data

Daily number of births, in Quebec, January 1, 1977 to December 31, 1990.



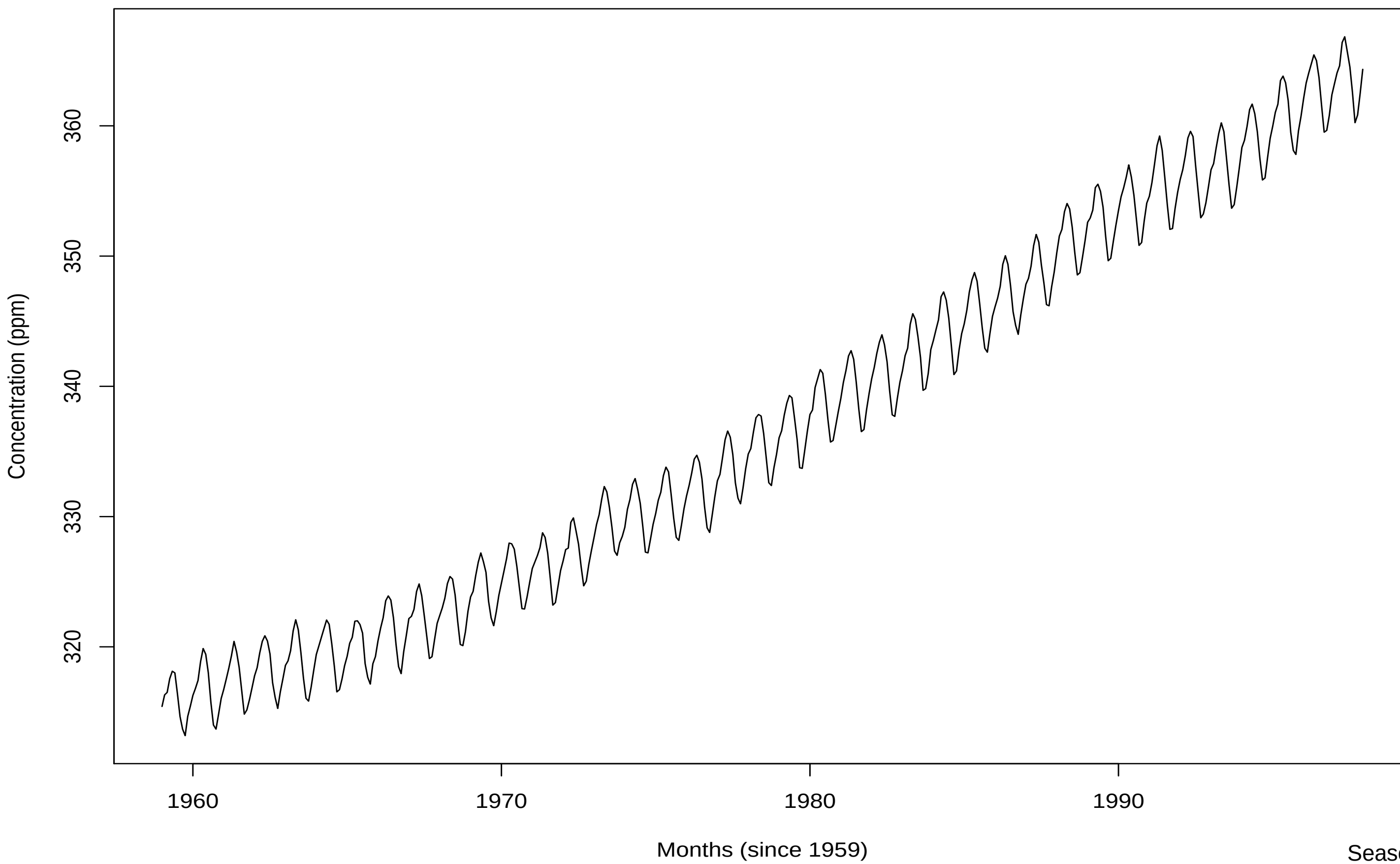
Example : Quebec Births data

Weekly aggregates (from R.J. Hyndman's *Time Series Data Library*, `http://www-personal.buseco.monash.edu.au/`)



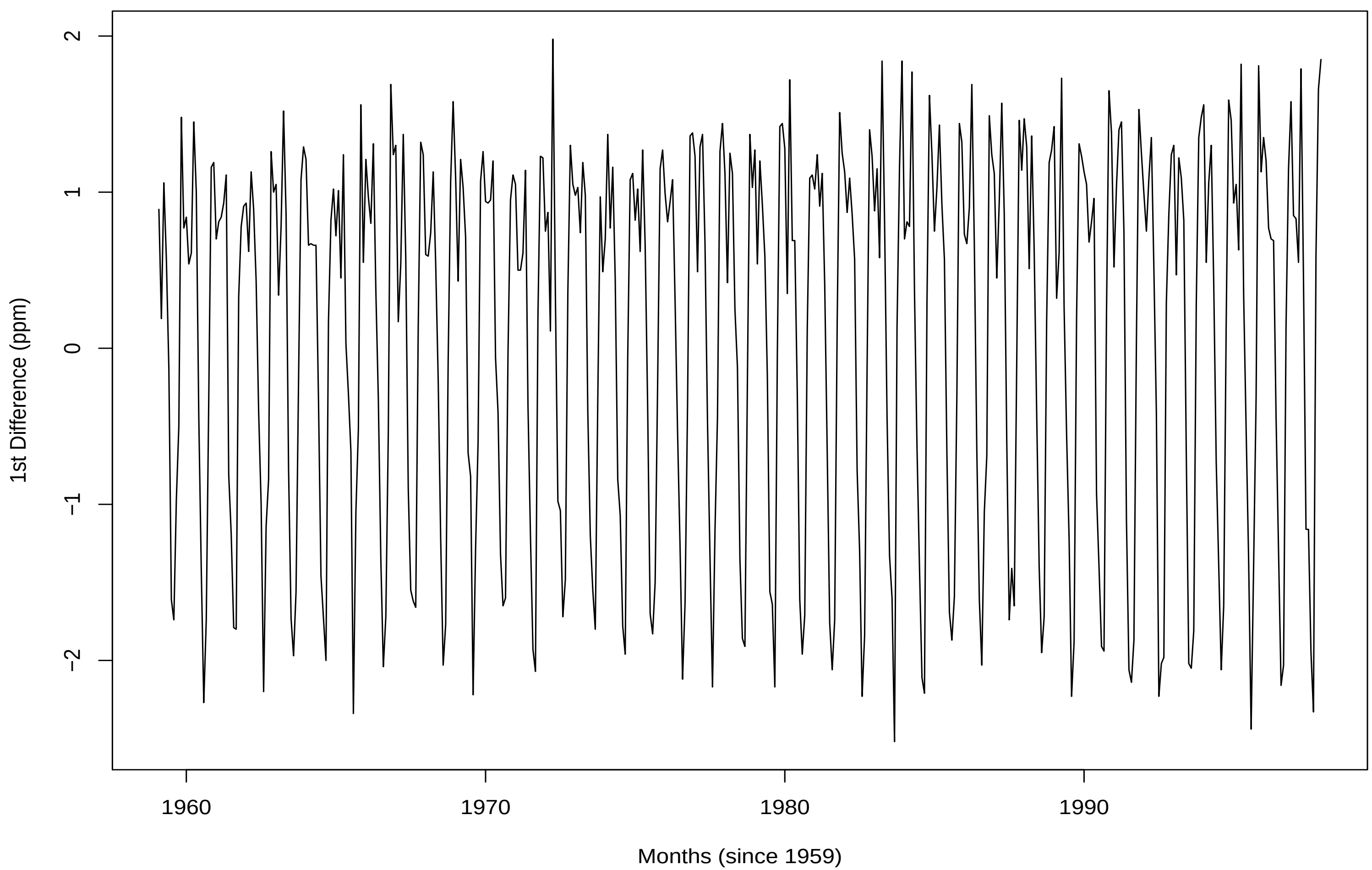
Example : CO_2 data

468 monthly observations atmospheric concentrations of CO_2 from 1959 to 1997 (from R)



Example : CO_2 data

First differences (not the only transformation possible ...)



Modelling Context

- Discrete time process $\{X_t\}$, zero mean
- Time series data $\{x_t : t = 1, \dots, N\}$
- Parametric models, Gaussian residual errors
- Stationarity (after pre-processing)
- require
  - likelihood-based inference about system parameters,
  - prediction/forecasting.

Need

- likelihood,
- inference mechanism.

Time Domain modelling

Most commonly, but not exclusively, data collected in the time domain

- finance, banking, econometrics
- climatology
- earth sciences
- computing systems, internet traffic, pageview statistics etc.

Natural to seek inferences and make predictions in the time domain.

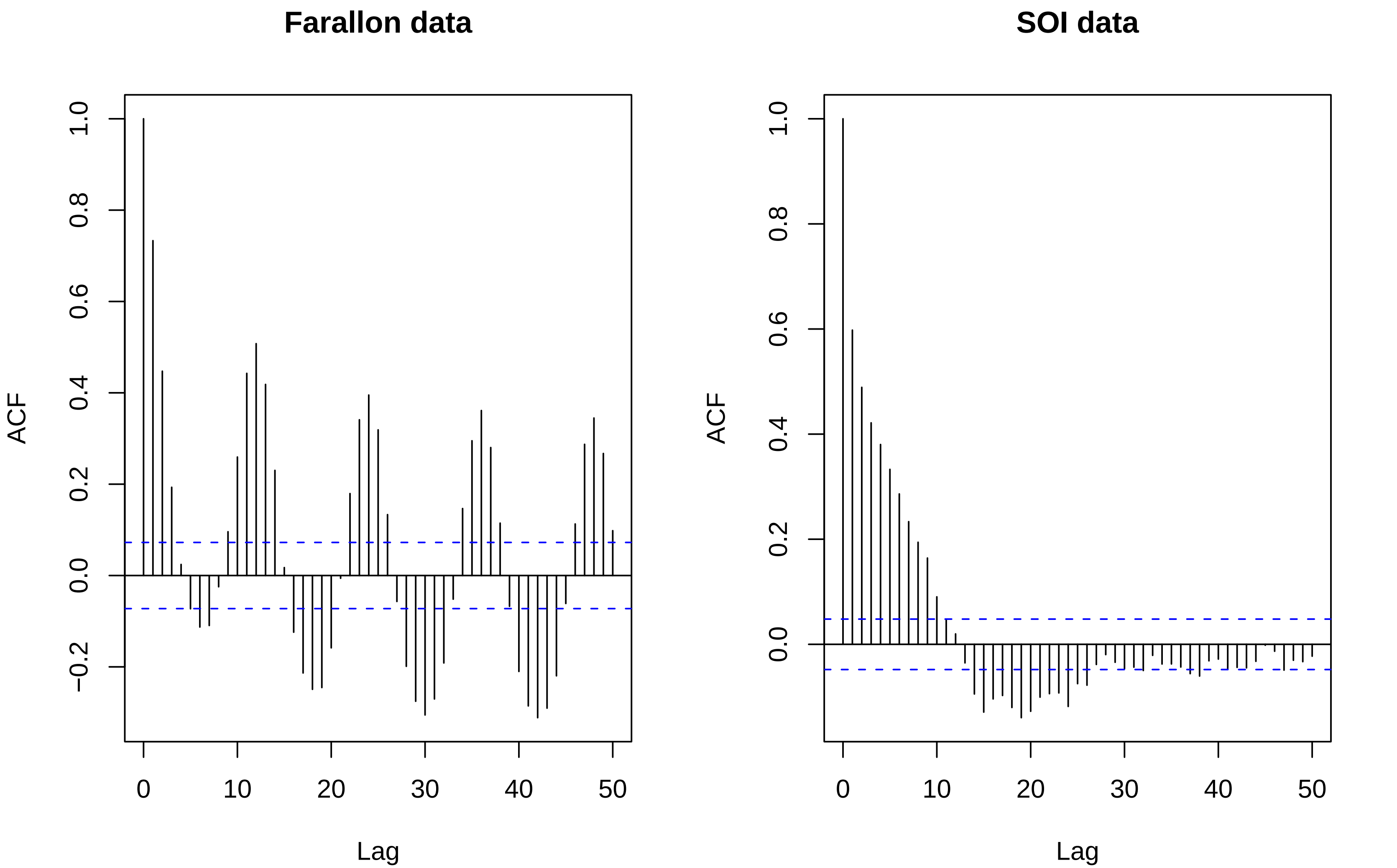
Autocovariance

In the time domain, raw data plots are not that informative; the stationary process is characterized by its autocovariance sequence (acvs)

$$\gamma_k = \mathsf{E}[X_t X_{t+k}] \qquad k \in \mathbb{Z}$$

Determination of underlying structure is more straightforward after inspection of the acvs/autocorrelation sequence.

ACF plot comparison



Starting Point: ARIMA Models

$\{X_t\}$ with zero mean follows an ARIMA(p, d, q) model if

$$\Phi(B)\Delta^d X_t = \Theta(B)\epsilon_t,$$

- B is the backward shift operator, so that $BX_t = X_{t-1}$.
- $\Delta^d = (1 - B)^d$, where d is the positive, integer-valued level of differencing required to achieve stationarity.
- $\{\epsilon_t\}$ is a zero mean Gaussian error process with variance σ_ϵ^2.
- yields a Gaussian likelihood in the time domain.

The process is specified by the AR and MA polynomials

$$\Phi(B) = 1 - \phi_1 B - ... - \phi_p B^p \qquad \Theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$$

with real-valued parameters $\phi_j \neq 0, \theta_j \neq 0$.

Sufficient conditions for *stationarity* and *invertibility* of the differenced series $Y_t = \Delta^d X_t$ are that the roots of the two polynomial equations

$$\Phi(z) = 0 \qquad \Theta(z) = 0$$

lie outside the unit circle. Under these conditions, ARMA processes have infinite AR and MA representations.

Parameterization

Under the restrictions of stationarity/invertibility, the parameter space for general (p, q) is a region in $\mathbb{R}^{p+q}$ that is not straightforward.

This makes inference (e.g. MCMC) difficult to implement.

Can use a reciprocal root parameterization (Huerta and West (1999)): write

$$\Phi(z) = \prod_{j=1}^{p}(1 - \zeta_{1j} z) \qquad \Theta(z) = \prod_{j=1}^{q}(1 - \zeta_{2j} z),$$

where $(\zeta_{11}, ..., \zeta_{1p})$ and $(\zeta_{21}, ..., \zeta_{2q})$ are complex-valued parameters.

For both Φ and Θ, the reciprocal roots are either real, or appear in complex conjugate pairs. Denote by

- (p_1, p_2), and (q_1, q_2) the number of real and complex conjugate pairs (so that $p = p_1 + 2p_2$, $q = q_1 + 2q_2$).
- For the complex roots, for $r = 1, 2$

$$\begin{aligned} \zeta_{rj} &= \alpha_{rj} e^{i2\pi\omega_{rj}} &= \alpha_{rj} \cos 2\pi\omega_{rj} + i\alpha_{rj} \sin 2\pi\omega_{rj}; \\ \zeta_{rj+1} &= \alpha_{rj} e^{-i2\pi\omega_{rj}} &= \alpha_{rj} \cos 2\pi\omega_{rj} - i\alpha_{rj} \sin 2\pi\omega_{rj}. \end{aligned}$$

- For stationarity/invertibility:

$$0 \leq \alpha_{1j}, \alpha_{2j} < 1 \quad 0 \leq \omega_{1j}, \omega_{2j} < 1/2.$$

MCMC Inference in the Time Domain

Using this parameterization, it is possible to implement (transdimensional) MCMC in a reasonably straightforward fashion.

- Model space defined by p, q,
- Birth/Death moves to change dimension,
- Metropolis-Hastings moves on θ and ϕ,
- Can incorporate different degrees of differencing.

However **computation of Gaussian likelihood requires inversion of potentially large covariance matrix.** This can be prohibitive.

Frequency Domain Representation

Spectral representation of the acvs of $\{X_t\}$:

$$\gamma_k = \int_{-1/2}^{1/2} \exp\{2\pi i k f\}\, dS_I(f) = \int_{-1/2}^{1/2} \exp\{2\pi i f k\}\, S(f) df.$$

- $S_I(f)$ is a non-decreasing function on $(-1/2, 1/2)$,
- derivative $S(f)$, the Spectral Density Function (SDF)
- $S(f)$ gives a decomposition of the variance of the process into contributions of different frequencies.

$$S(f) = \sum_{k=-\infty}^{\infty} \gamma_k \exp\{-2\pi i f k\}.$$

SDF for ARMA processes

The stationary $ARMA(p,q)$ process $\{X_t\}$ has SDF

$$S(f) = \sigma_\epsilon^2 \frac{|\Theta(e^{i2\pi f})|^2}{|\Phi(e^{i2\pi f})|^2} = \sigma_\epsilon^2 \frac{\prod_{j=1}^{q} \left|1 - \zeta_{2j} e^{i2\pi f}\right|^2}{\prod_{j=1}^{p} \left|1 - \zeta_{1j} e^{i2\pi f}\right|^2}.$$

This function is

- bounded
- bounded away from zero.

The Periodogram

At the Fourier frequencies, $f_j = j/N$, $j = 0, \ldots, N/2$, define periodogram, I, by

$$I(f_j) = |Z(f_j)|^2 \quad j = 0, \ldots, N/2$$

where Z is the Discrete Fourier Transform (DFT) of $\{X_t\}$

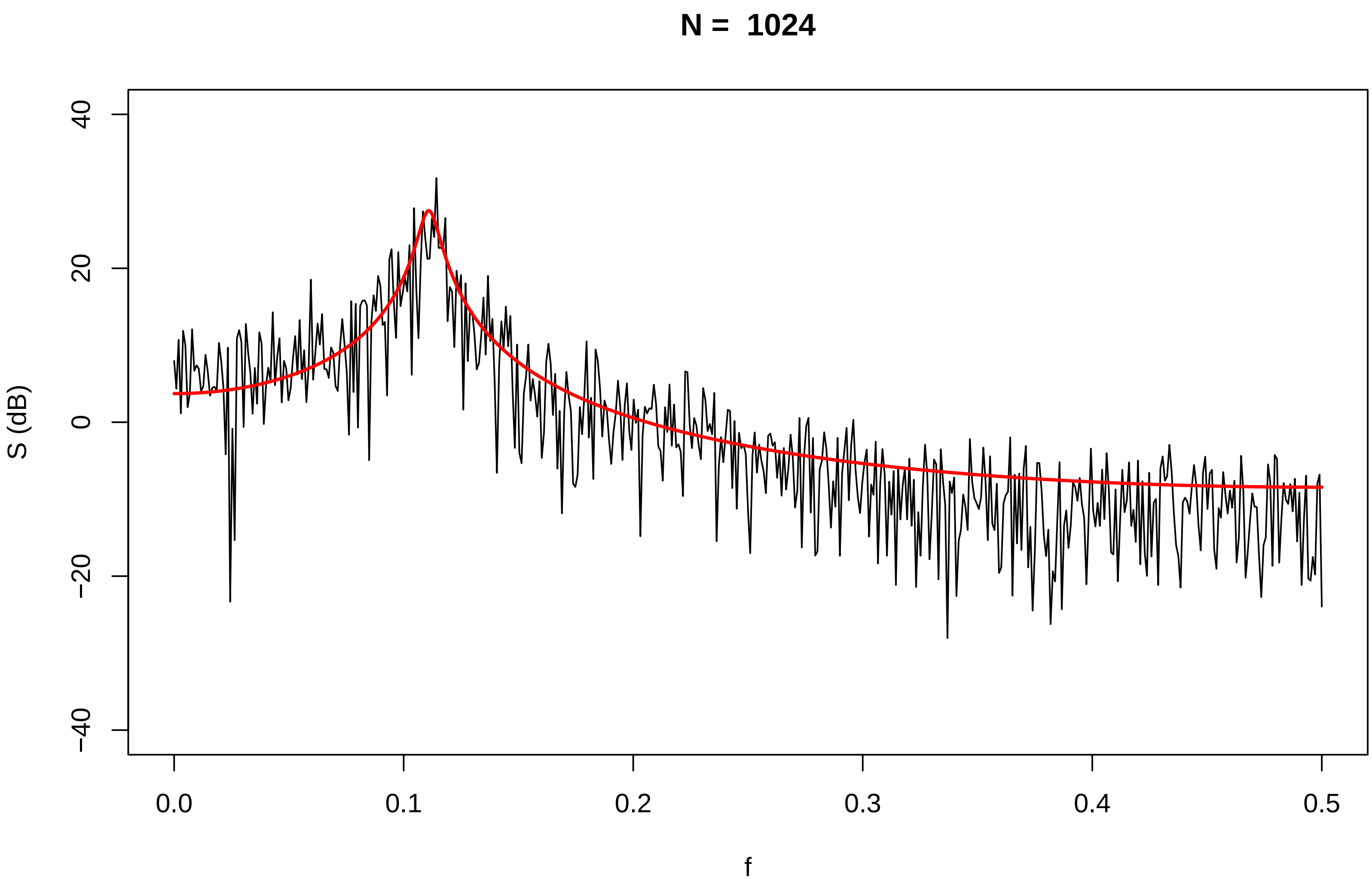
$$\begin{aligned} Z(f_j) &= \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-i2\pi t f_j} = A_j + iB_j \\ I(f_j) &= \frac{1}{N} \left[\sum_{t=0}^{N-1} X_t^2 + 2 \sum_{t=1}^{N-1} \sum_{s=0}^{t-1} X_t X_s \cos\left(2\pi j (t-s)/N\right) \right] \end{aligned}$$

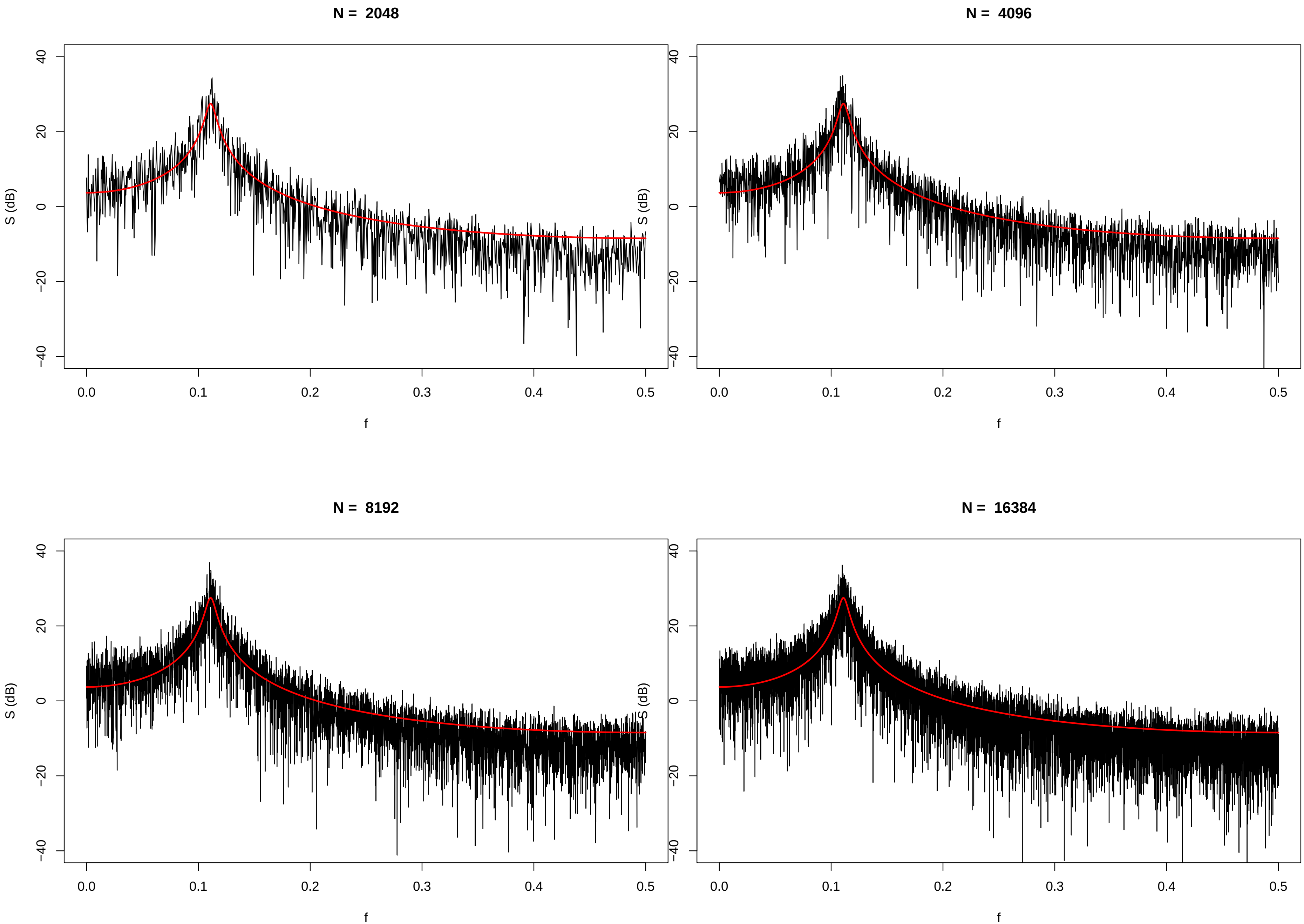
The periodogram is a natural non-parametric estimator for S

- $I(f_j)$ and $I(f_k)$, $j \neq k$ are asymptotically independent
- I asymptotically unbiased for S (in most cases)
- inconsistent
- always finite valued.

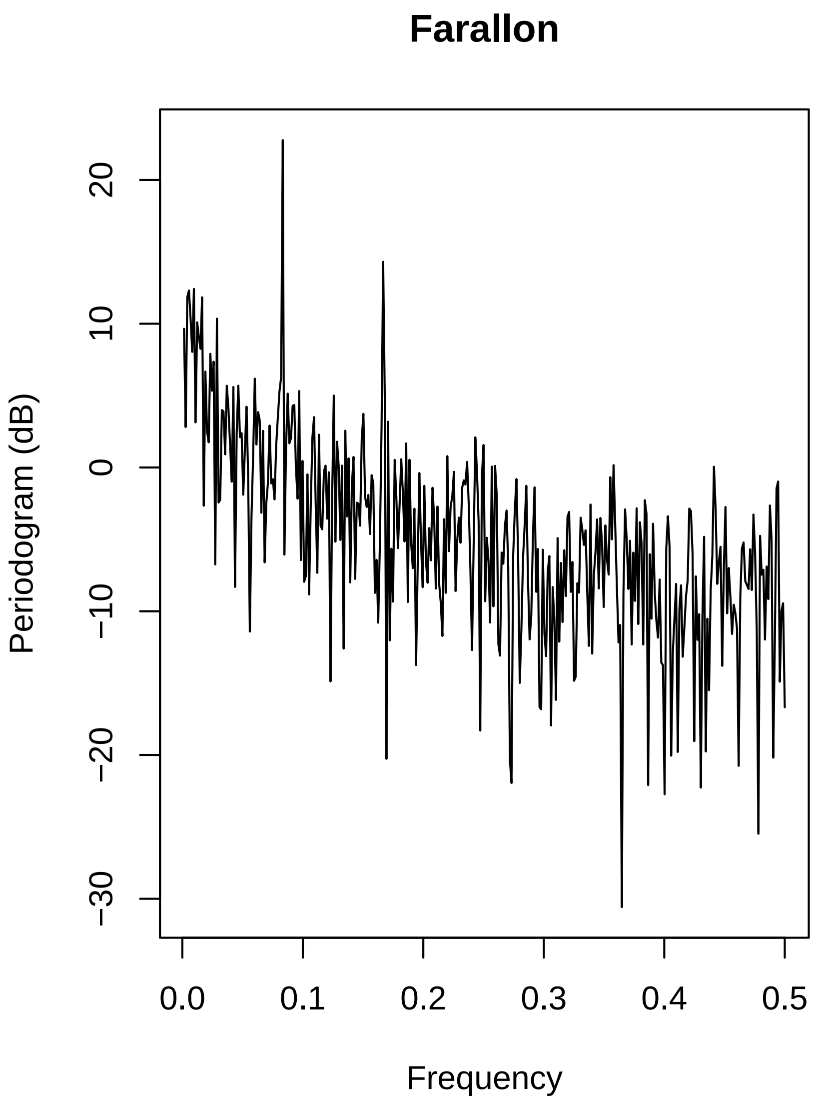
Behaviour as $N \to \infty$ illustrates problem; the grid of Fourier frequencies has $O(N)$ components, so the learning rate is zero (we essentially have just a data transformation).

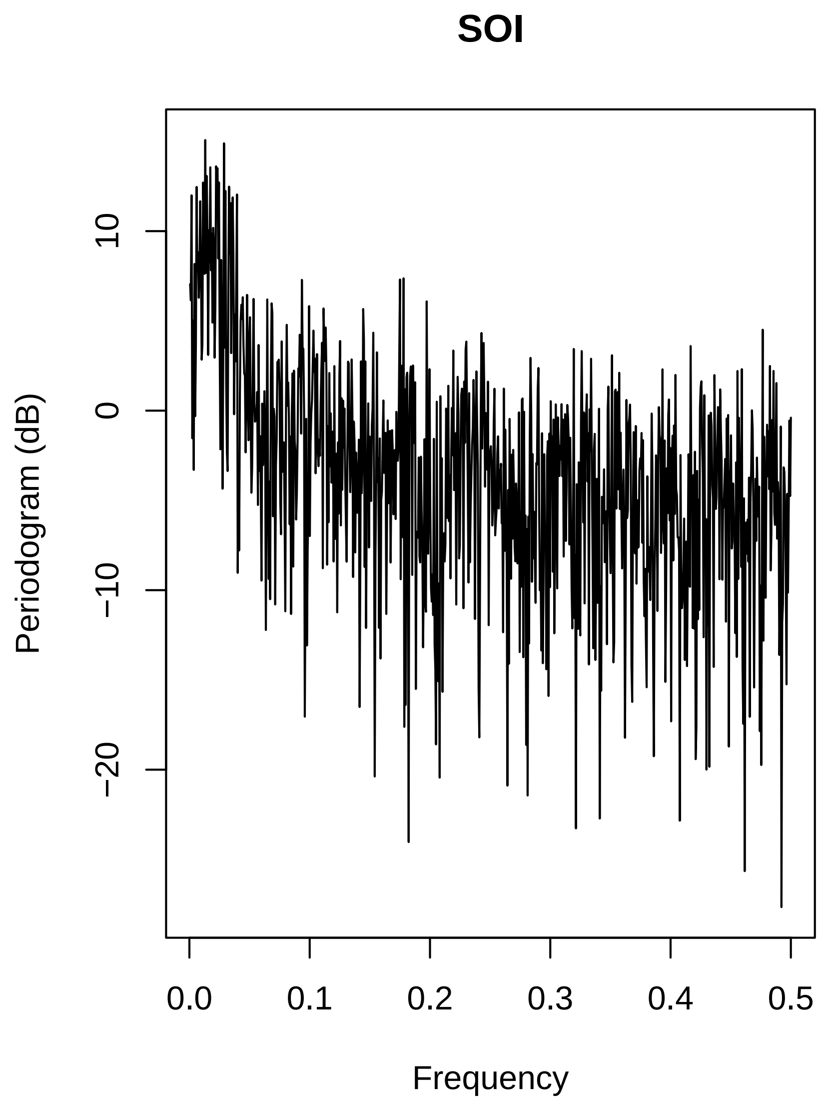
$N = 1024$



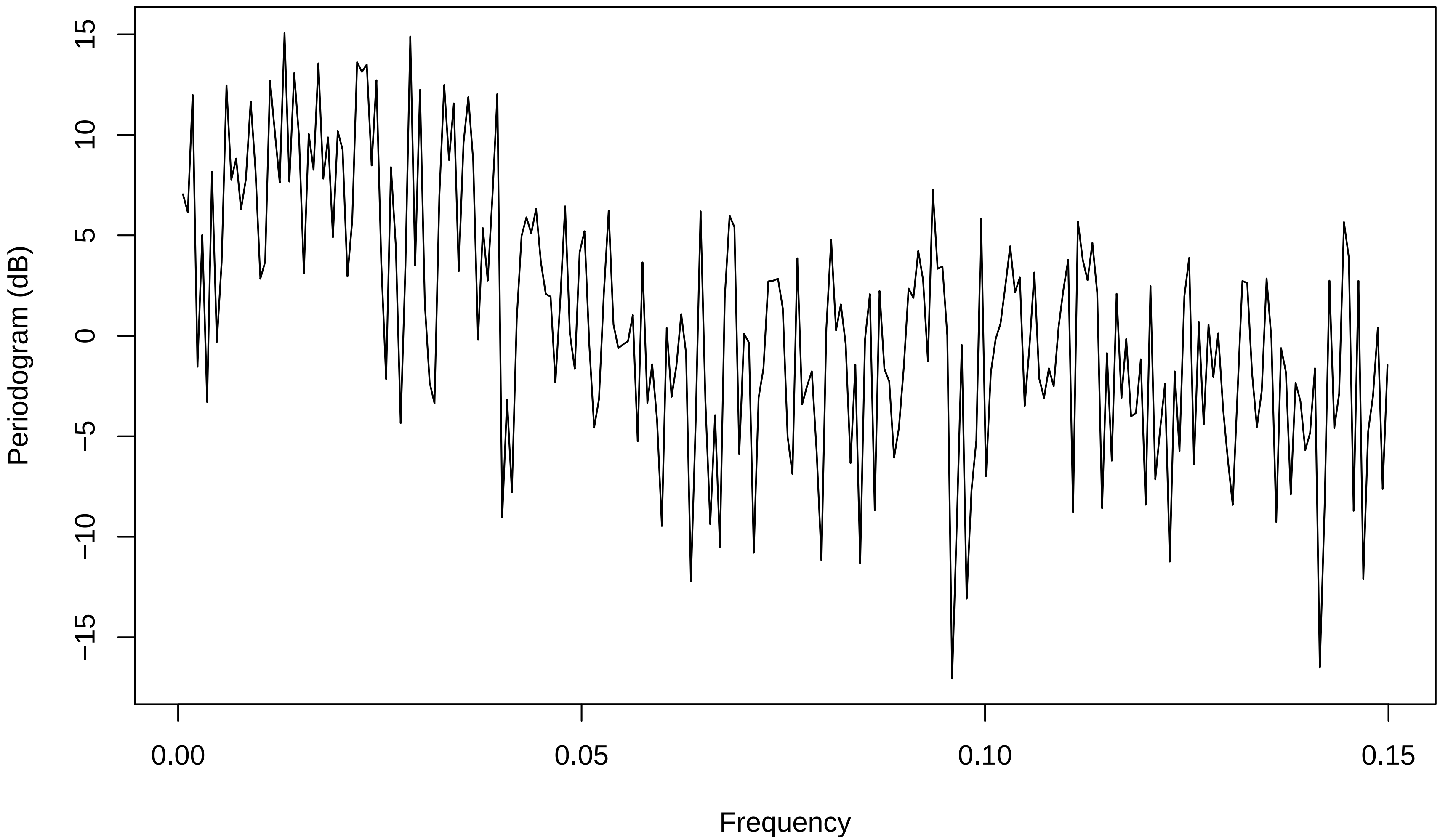

N = 2048
N = 4096
N = 8192
N = 16384
S (dB)
f
40
20
0
−20
−40
0.0
0.1
0.2
0.3
0.4
0.5

Examples Revisited





SOI periodogram



Likelihood Inference

See Chan and Palma (2004) for a recent summary:

- exact time domain
- approximate time domain
  - AR approximation
  - MA approximation
  - QML
- exact frequency domain
- approximate frequency domain

Most exact methods are at least $O(N^2)$ per likelihood evaluation in computation.

The Whittle Likelihood

Motivation: Approximate the Gaussian time-domain likelihood using a spectral approximation to covariance matrix Σ (Whittle (1951, 1953), Grenander and Szëgo (1958))

$$\begin{aligned}\frac{1}{N}\log L(\theta,\phi) &= -\frac{1}{2N}\log|\Sigma(\theta,\phi)| - \frac{1}{2N}X^T\Sigma(\theta,\phi)^{-1}X \\ &\approx -\frac{1}{2N}\left\{\sum_{j=1}^{N}\log S(\omega_j;\theta,\phi) + \sum_{j=1}^{N}\frac{I(\omega_j)}{S(\omega_j;\theta,\phi)}\right\}\end{aligned}$$

where $\omega_j = 2\pi f_j = 2\pi j/N$.

Statistical Properties

For models with bounded spectra, for N large,

$$I(f_j)/S(j/N) \stackrel{\mathfrak{L}}{=} U_j, \qquad j = 0, \ldots, M = N/2$$

- $U_j \sim \chi_2^2/2 \equiv Exp(1)$ i.i.d., $j = 1, \ldots, M-1$,
- $U_0, U_M \sim \chi_1^2$

These properties facilitate likelihood-based inference via the Whittle approximation.

ML estimates of SDF parameters are consistent and asymptotically normally distributed.

Statistical Properties

For models with bounded spectra, for N large,

$$I(f_j)/S(j/N) \stackrel{\mathfrak{L}}{=} U_j, \qquad j = 0, \ldots, M = N/2$$

- $U_j \sim \chi_2^2/2 \equiv Exp(1)$ i.i.d., $j = 1, \ldots, M-1$,
- $U_0, U_M \sim \chi_1^2$

These properties facilitate likelihood-based inference via the Whittle approximation.

ML estimates of SDF parameters are consistent and asymptotically normally distributed.

Seasonally Persistent Processes

It is plausible in many contexts that $\{X_t\}$ has a seasonal component

- calendar-based data collection (annual, quarterly, monthly, or weekly cycles)
- high dependence at specific lags
- using seasonal differencing e.g. for monthly data

$$Y_t = (1 - B^{12})X_t = X_t - X_{t-12}$$

removes an annual seasonal component.

Seasonal processes are not stationary before differencing, and have SDFs with singularities (*poles*) that render the SDF not integrable.

Persistence: A process $\{X_t\}$ with acvs $\{\gamma_k\}$ can also exhibit persistence, that is,

- long-range dependence if, $\forall a > 0$

$$\lim_{k \to \infty} \frac{a^{-k}}{\gamma_k} = 0$$

that is, the acf is slowly decaying.

- long-memory if the acvs is absolutely divergent

$$\sum_k |\gamma_k| = \infty$$

- in practice, diagnosed by observing large autocorrelation at high lags, spectral power near frequency zero.

Constructing Persistent Processes

Let $\{W_t\}$ be an i.i.d. Gaussian sequence with variance 1. Let $\delta \in (-1/2, 1/2)$, and write

$$(1-B)^{\delta} = \sum_{k=0}^{\infty} c_k(-\delta)(-B)^k \qquad c_k(d) = \frac{\Gamma(k+d)}{\Gamma(k+1)\Gamma(d)}$$

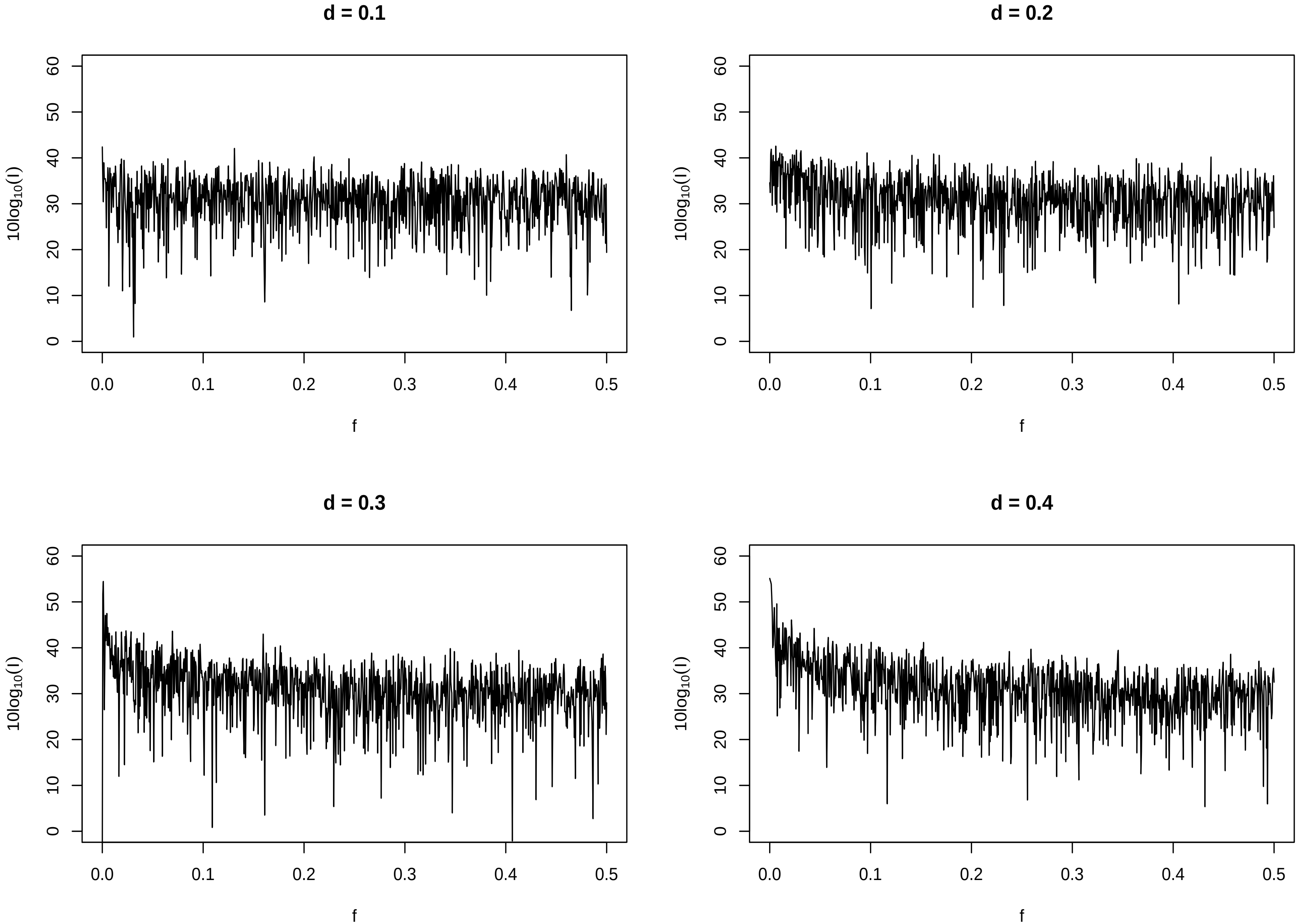
and set $X_t = (1-B)^{-\delta} W_t$.

This *fractional differencing* yields a process that is stationary if $\delta < 1/2$, long-memory if $0 < \delta < 1/2$ and long-range dependent if $-1/2 < \delta < 1/2$. For k large,

$$\gamma_k \sim k^{-(1-2\delta)}$$

The persistence is *associated with frequency zero*.

Periodograms for different δ.



Seasonal Persistence

Similar construction: replace $\{c_k\}$ sequence by $\{g_k\}$ such that, for some $\lambda_0 \in (0, 1/2)$,

$$X_t = (1 - 2\cos(2\pi\lambda_0)B + B^2)^{-\delta} W_t$$

Recursion for $\{g_k\}$ given by $g_{-1} = 0, g_0 = 1$ and for $k > 0$

$$g_k = \left(\frac{2}{k+1}\right)(\delta + k)\cos(2\pi\lambda_0) - \left(\frac{2\delta + k - 1}{k+1}\right) g_{k-1}$$

but no simple explicit form.

$\{g_k\}$ are coefficients of the *Gegenbauer* polynomials (see Gray, Zhang, Woodward (1989), Lapsa(1997)).

This procedure yields a process $\{X_t\}$ that has persistence associated with the frequency λ_0, and is stationary

- if $\delta < 1/2$ when $\lambda_0 \neq 0$, or
- if $\delta < 1/4$ when $\lambda_0 = 0$

SDF has relatively straightforward form

$$S(f) = \frac{1}{\left(2\left|\cos(2\pi f) - \cos(2\pi\lambda_0)\right|\right)^{2\delta}}$$

with

$$S(f) \to \frac{1}{\left(2\left|\sin(2\pi\lambda_0)\right|\right)^{2\delta}} \frac{1}{\left|2\pi f - 2\pi\lambda_0\right|^{2\delta}} \qquad f \to \lambda_0$$

ACV/ACF less straightforward

$$\begin{aligned}\sigma_X^2\gamma_k &= \frac{\Gamma(1-2\delta)}{\sqrt{\pi}2^{1/2+2\delta}}\left\{\sin(2\pi\lambda_0)\right\}^{1/2-2\delta} \\ &\quad \left[P_{k-1/2}^{2\delta-1/2}(\cos(2\pi\lambda_0)) + (-1)^k P_{k-1/2}^{2\delta-1/2}(-\cos(2\pi\lambda_0))\right]\end{aligned}$$

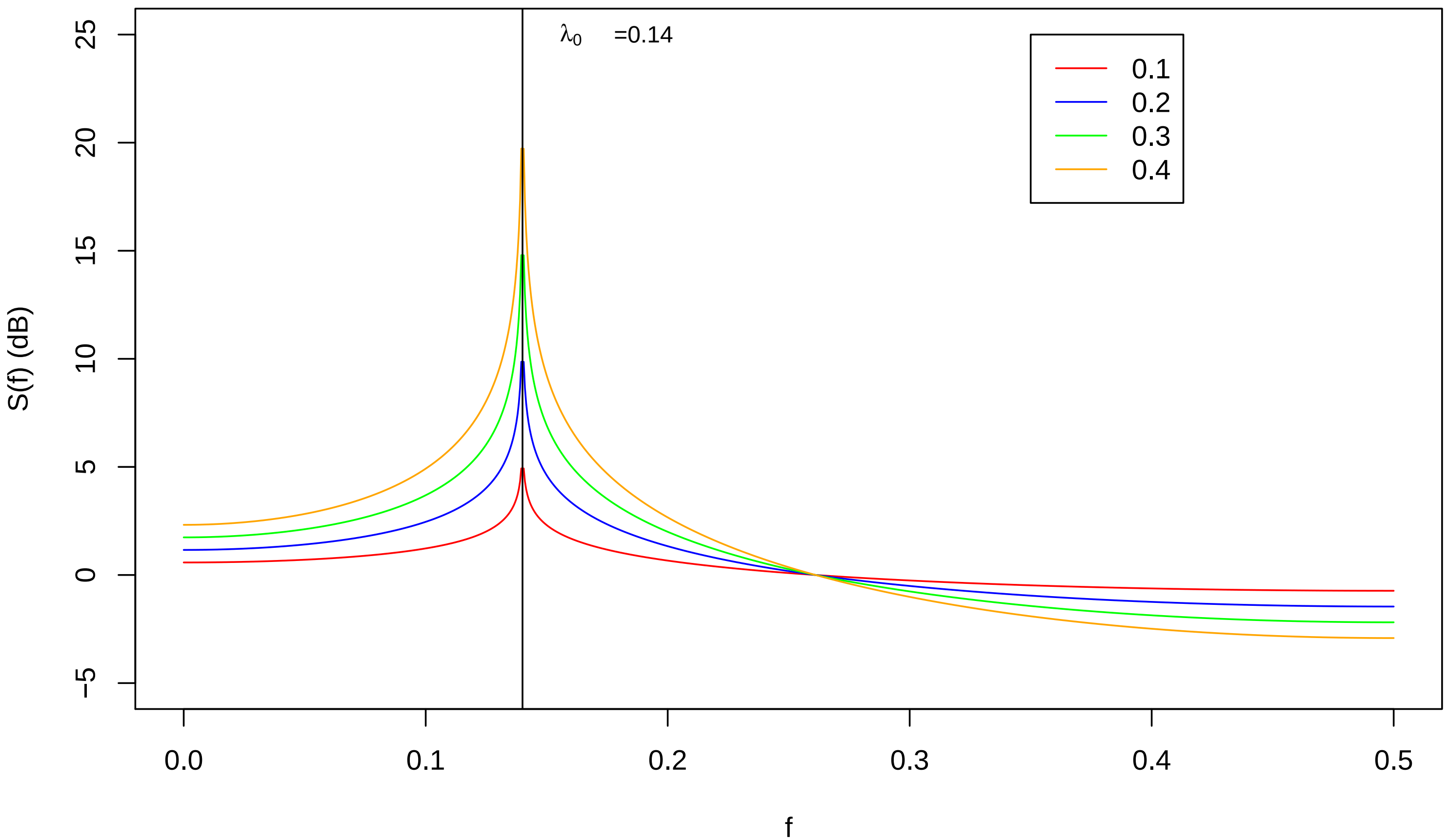
where $P_\nu^\mu(x)$ is the associated Legendre function of the first kind.

A recursion formula for $P_\nu^\mu(x)$ gives the acvs to arbitrary lag.

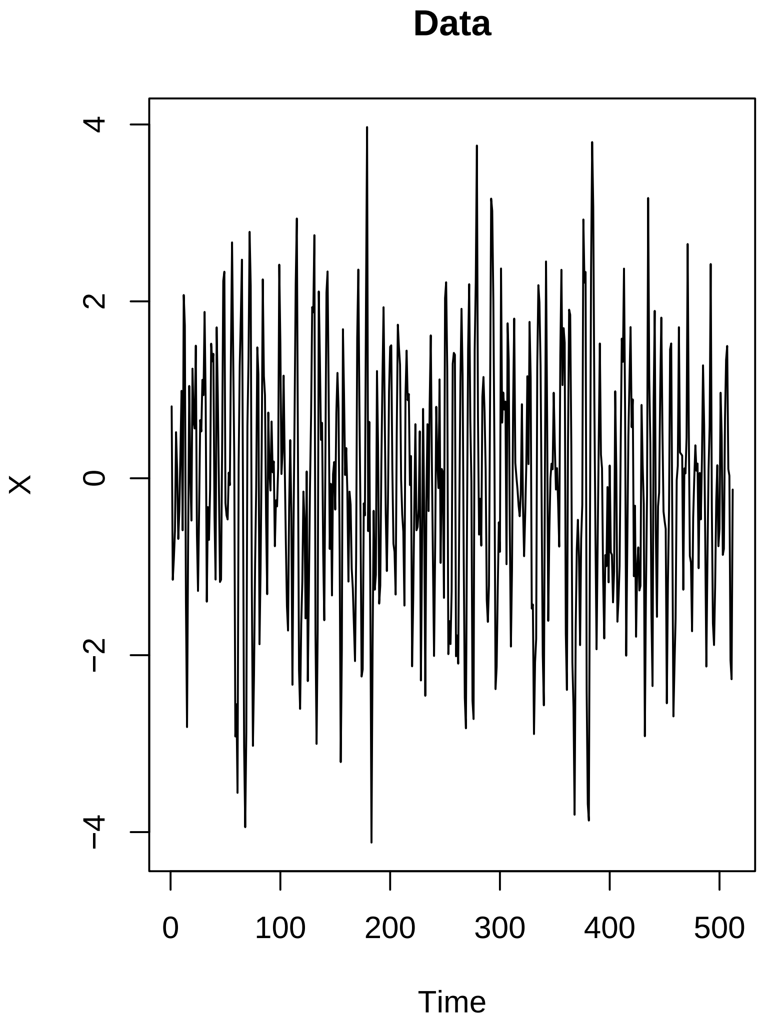
$$(\nu-\mu+1)P_{\nu+1}^\mu(x) = (2\nu+1)P_\nu^\mu(x) - (\nu+\mu)P_{\nu-1}^\mu(x)$$

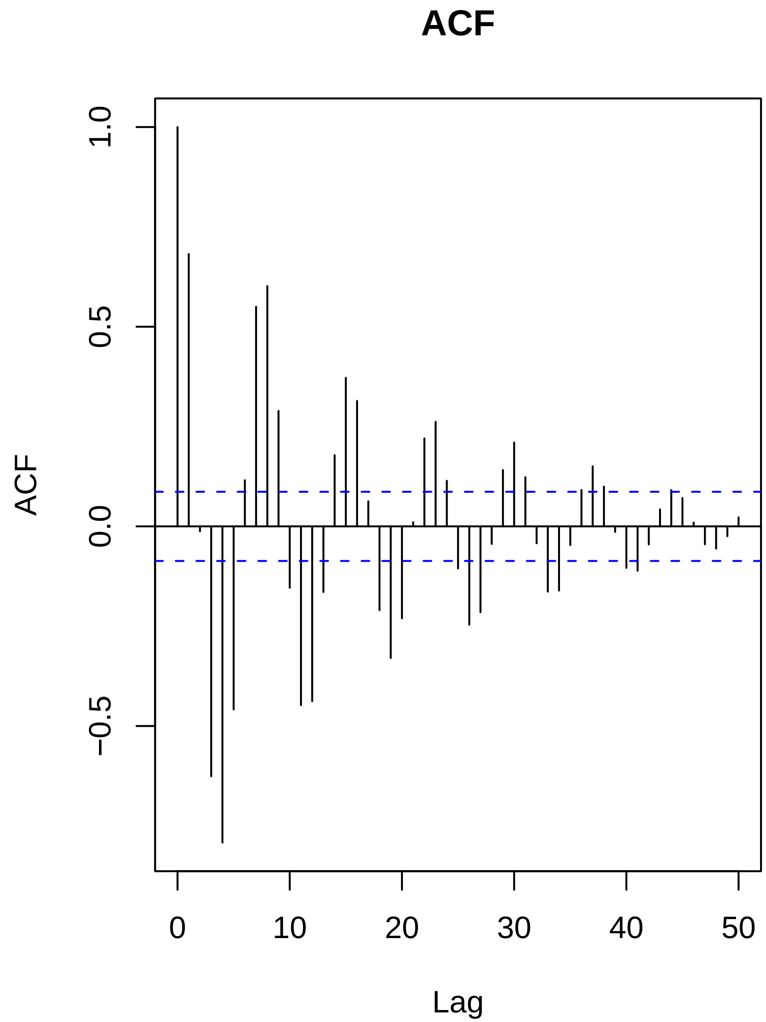
Gegenbauer Models

Characteristic singularity (pole) in the spectrum at λ_0.



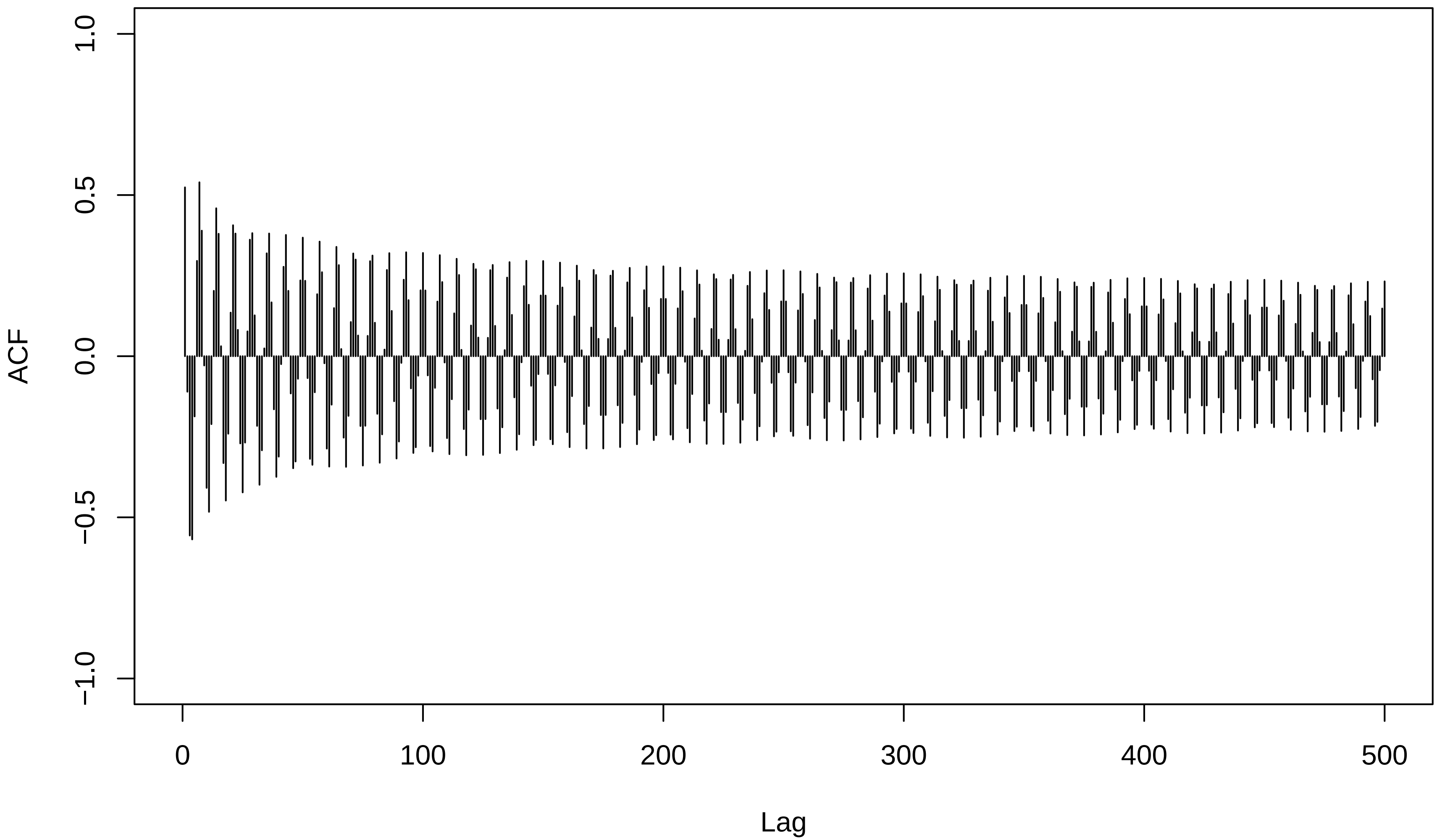
Example: $\lambda_0 = 0.14$, $\delta = 0.4$



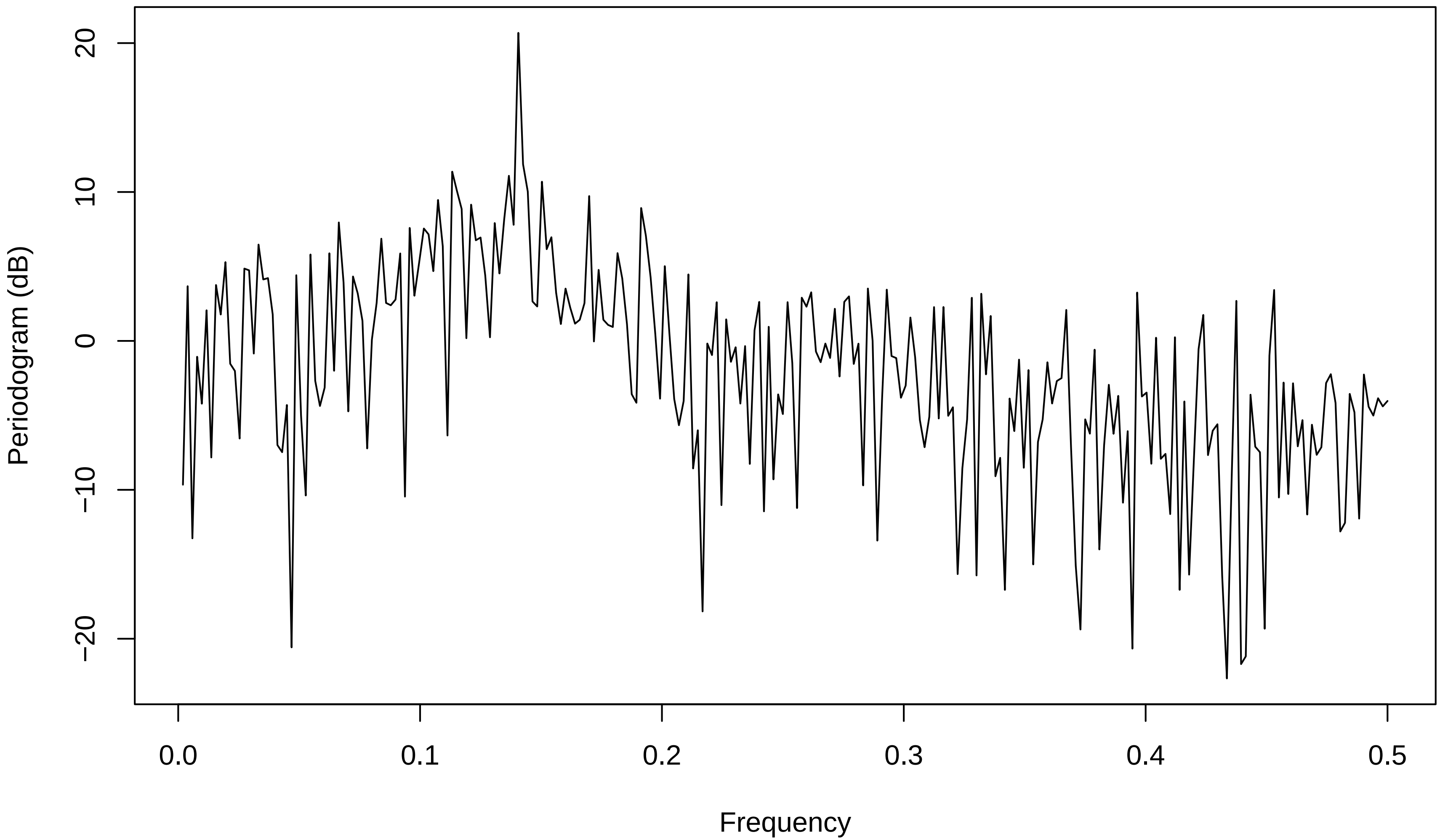


Theoretical ACF

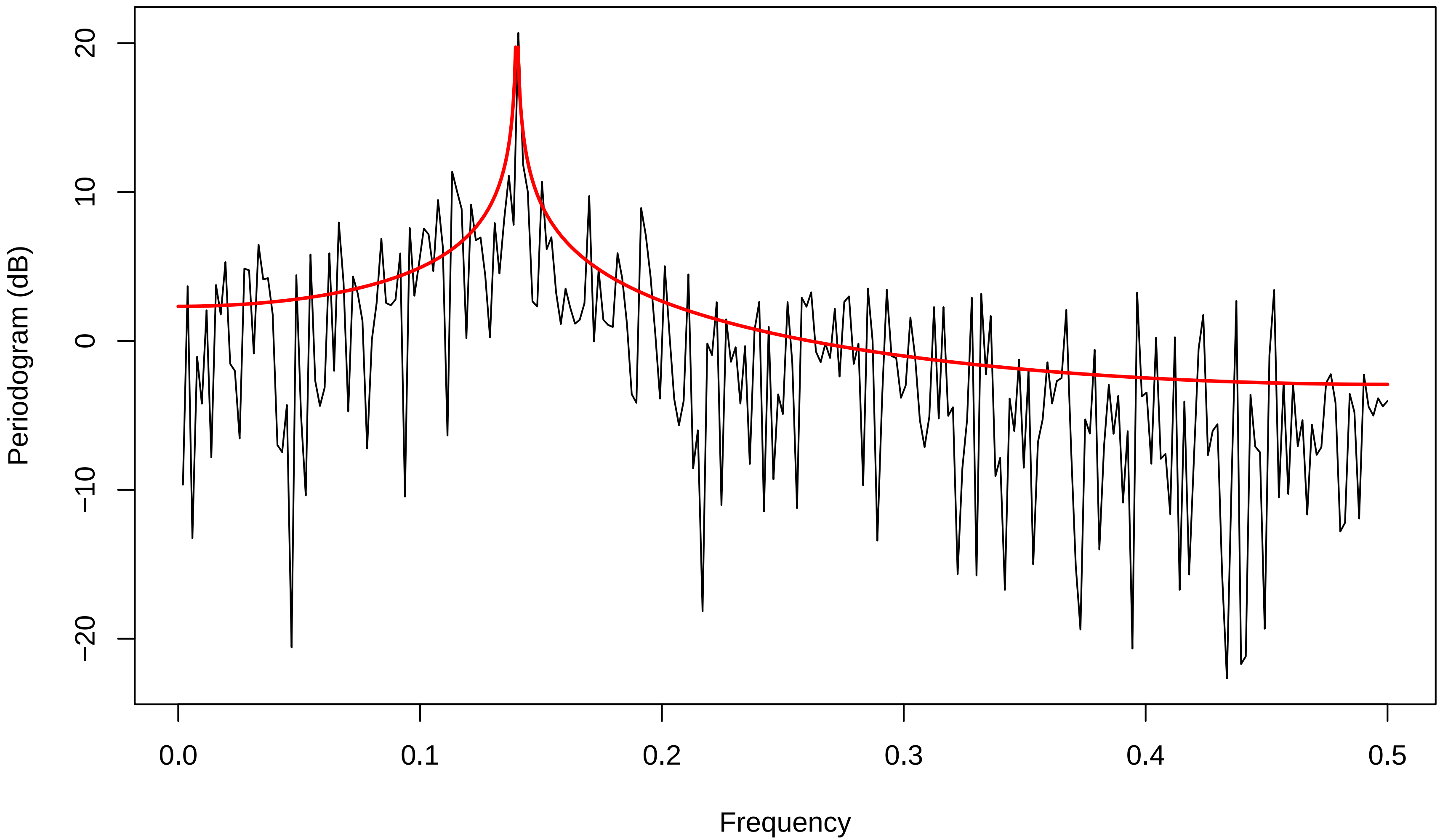
With $\delta = 0.4$, large autocorrelation at high lag separation.



Periodogram



Periodogram and SDF



The Whittle Likelihood for SP process

The distributional assumptions central to the construction of the Whittle likelihood **break down** for Fourier frequencies near to spectral poles.

Recall that the periodogram is finite at the pole λ_0*, where the SDF is not finite.*

When the pole is at $\lambda_0 = 0$ (standard long-memory), it is possible to deduce distributional properties for the periodogram evaluated at the Fourier frequencies.

For example, Hurvich and Beltrao (1993): when the pole is at $\lambda_0 = 0$.

- *the periodogram values should not treated as i.i.d exponential random variables*
- *asymptotically, the periodogram values are distributed a weighted combination of two independent χ_1^2 random variables.*
- *the asymptotic relative bias in the periodogram is positive for most values of the δ parameter.*

We attempt to re-evaluate these results for general λ_0.

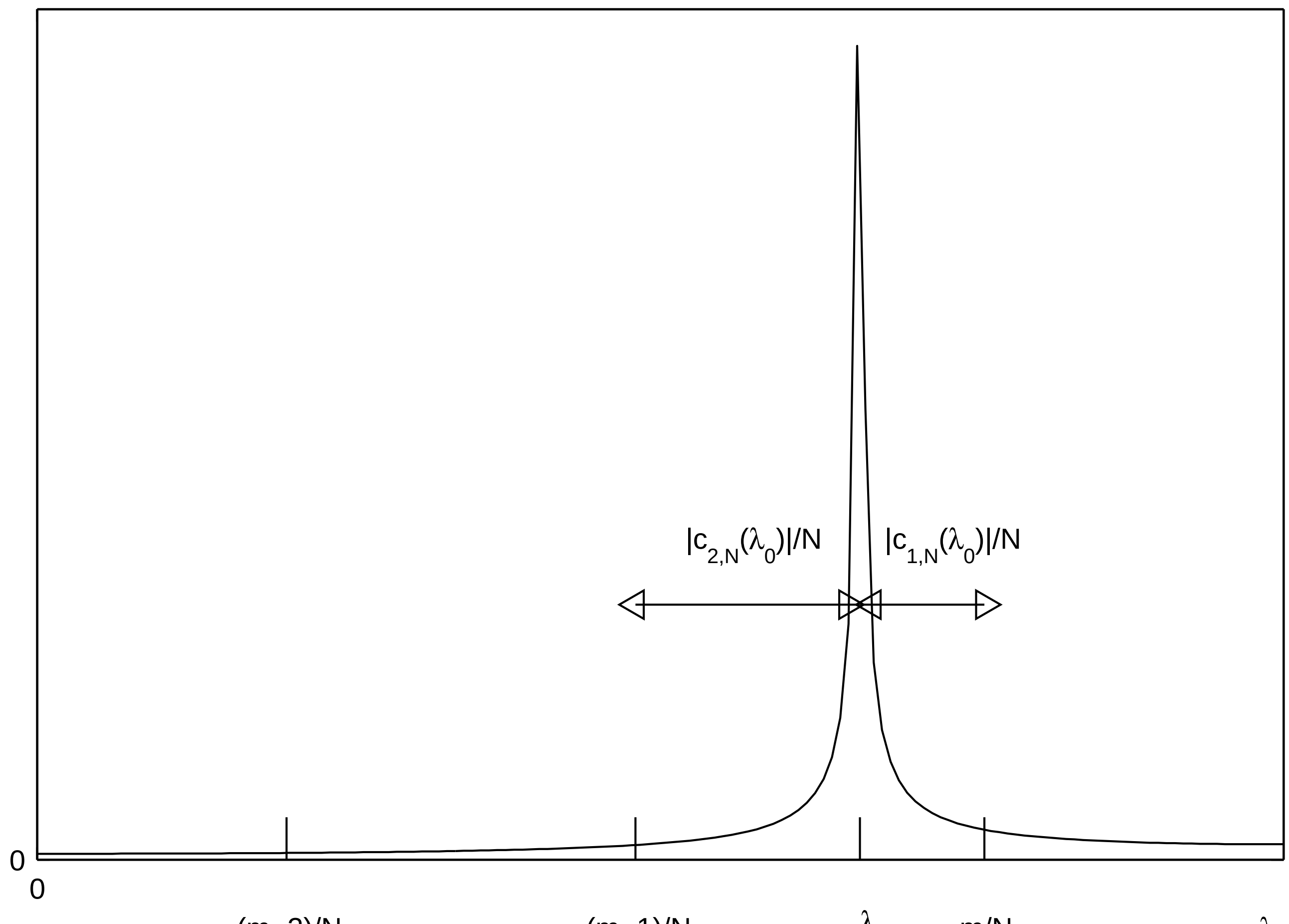
Relabelling the Fourier frequencies

We attempt to quantify the (relative) bias in the periodogram at Fourier frequencies near to λ_0.

- The bias will depend on the distance between the Fourier frequency and λ_0.
- For any fixed λ_0, the distance from the pole to the nearest Fourier frequency depends on sample size N.
- Intuitively, as the distance decreases, the bias increases.

For convenience, we re-label the Fourier frequencies, so that f'_j is the j^{th} closest Frequency to the pole.

The grid of Fourier frequencies m/N for integer m, and their relation to the Fourier frequencies closest to the pole. Here $\omega_1/(2\pi) = m/N$ and $\omega_2/(2\pi) = (m-1)/N$.



Quantifying the relative bias

The crucial factor determining the magnitude of the bias is the distance between the periodogram ordinates and λ_0.

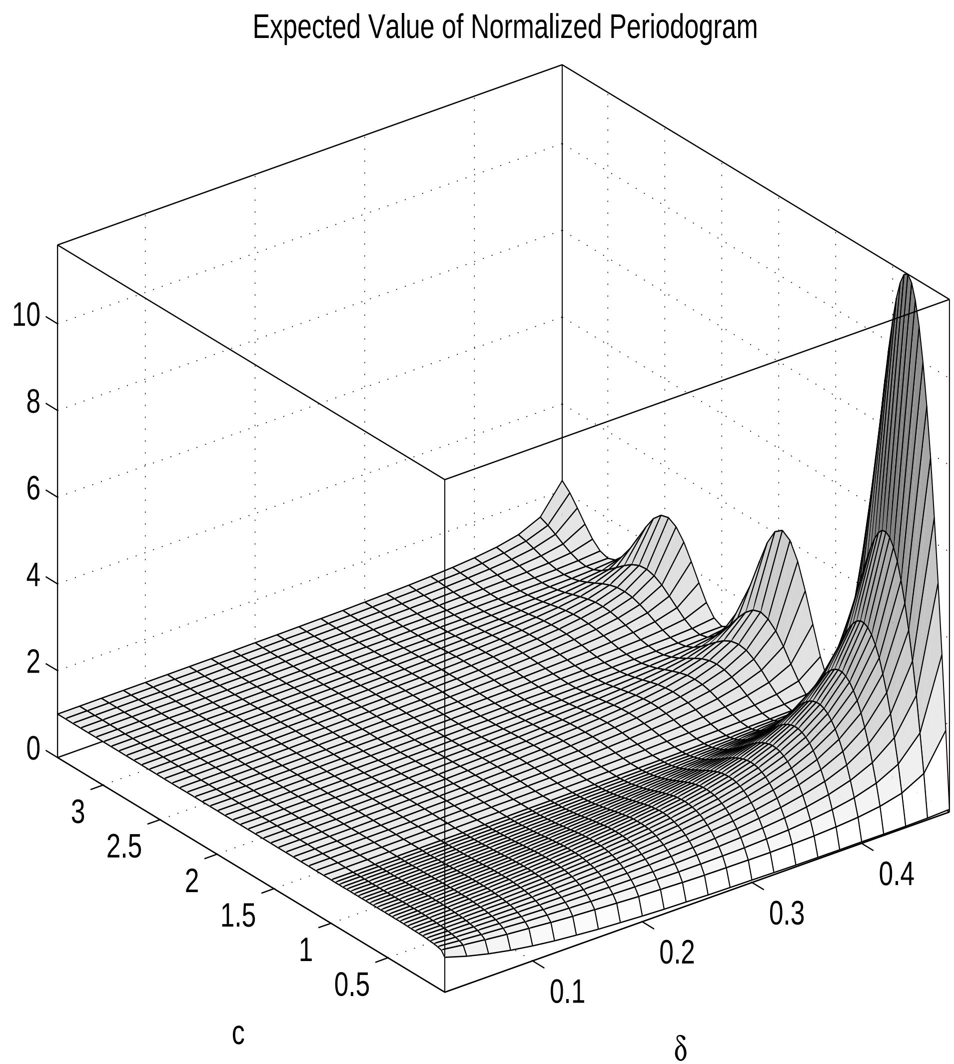
THEOREM : For (relabelled) Fourier frequency f_j', the j^{th} closest to the pole at λ_0, for large N

$$E\left(\frac{I\left(f_j'\right)}{S\left(f_j'\right)}\right) = \frac{2}{\pi}\int_{-\infty}^{\infty} \frac{\sin^2\left\{u/2 - \pi c_{j,N}(\lambda_0)\right\}}{\left\{u - 2\pi c_{j,N}(\lambda_0)\right\}^2} \left|\frac{2\pi c_{j,N}(\lambda_0)}{u}\right|^{2\delta} du$$

plus terms that are $o(1)$, where $c_{j,N}(\lambda_0) = j - N\lambda_0$.

Bias for various values of (c, δ)

The bias varies from minimal to considerable.



Demodulation

The Demodulated DFT (DDFT) or offset DFT (Pei and Ding (2004)) of X_t, with demodulation via frequency λ is denoted Z_λ, and is defined for $f_j = j/N$ by

$$Z_\lambda(f_j) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} X_t e^{-2i\pi(f_j+\lambda)t} = A_{\lambda,j} + iB_{\lambda,j}, \quad j = 0, \ldots, M.$$

The demodulated periodogram at frequency f_j with demodulation via λ is denoted $I_\lambda(f_j)$, and is defined via the ordinary periodogram I by

$$I_\lambda(f_j) = I(f_j + \lambda) = |Z_\lambda(f_j)|^2 = A^2_{\lambda,j}(j) + B^2_{\lambda,j}.$$

Demodulation simplifies the calculation of distributional properties of periodogram values near the spectral pole.

THEOREM : For a Gegenbauer (λ_0, δ) process with SDF

$$S(f) = S^{\dagger}(f)\,|f - \lambda_0|^{-2\delta}$$

where $S^{\dagger}$ is a bounded SDF, the expected value of the periodogram evaluated at the pole λ_0, after demodulation by λ_0, is

$$(2\pi N)^{2\delta}\{-2S^{\dagger}(\lambda_0)\Gamma(-1-2\delta)\}\cos\{\pi(1/2+\delta)\}\pi^{-1} + o(1)$$

which is $O(N^{2\delta})$.

A Whittle Likelihood Adjustment

Using the previous theorem, and demodulation, we can construct an adjusted Whittle likelihood to estimate the parameters in the SPP.

For a Gegenbauer model with parameters (λ_0, δ), construct a demodulated Whittle likelihood as follows:

- Compute the DDFT of sample data with demodulation via $\lambda_D = \lambda_0 - [N\lambda_0]/N$; this aligns a new Fourier grid precisely with λ_0.

- This yields periodogram values

$$I(\lambda_0 + J_1/N), \ldots I(\lambda_0 + J_2/N)$$

where $J_1 = -[N\lambda_0]$ and $J_2 = (N/2) - [N\lambda_0]$.

- Construct a likelihood under the model where

$$I(\lambda_0 + j/N) \sim Exp(\eta_j)$$

give independent contributions, and

$$\eta_j = \frac{|j|^{2\delta\chi\{j \neq 0\}}}{Q(\delta)^{\chi\{j=0\}} N^{2\delta} S^{\dagger}(\lambda_0 + j/N)}$$

where $Q(\delta) = -\Gamma\left(-1 - 2\delta\right)\cos\{\pi\left(1/2 + \delta\right)\}2^{2\delta+1}\pi^{2\delta-1}$.

This likelihood is bounded on $(0, 1/2) \times (0, 1/2)$.

It is not continuous in λ_0 due to the demodulation, but the discontinuities are $O(1)$, hence ignorable.

Numerical maximization yields ML estimates.

Theoretical properties of the estimators are not straightforward to establish.

In particular, the behaviour of the the estimator of λ_0 is non-standard.

Asymptotic Properties of the Estimators

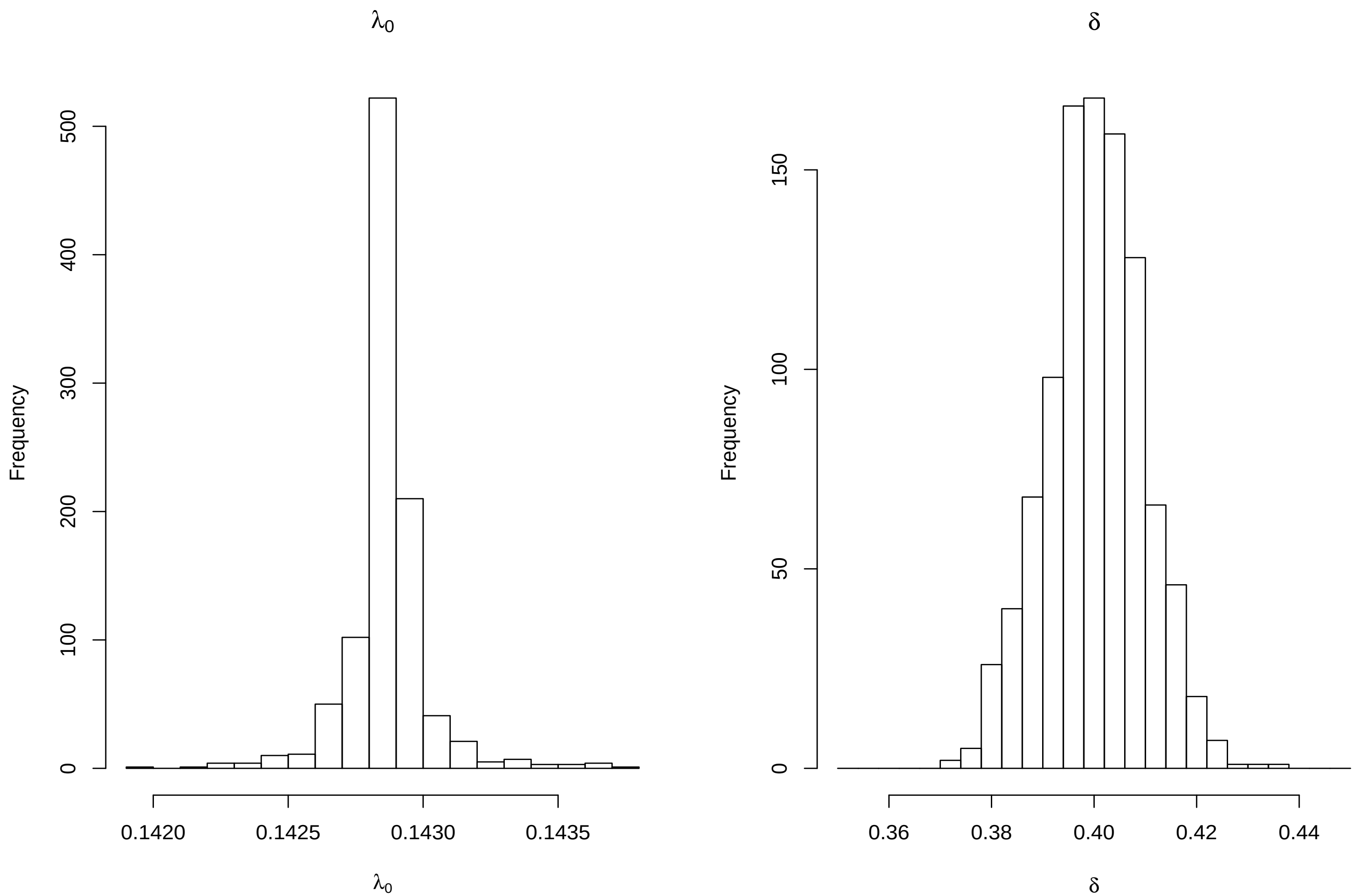
For the adjusted likelihood, we can establish

- N-consistency for $\hat{\lambda}_0$, $N^{1/2}$-consistency for $\hat{\delta}$
- asymptotic normality of $\hat{\delta}$
- asymptotic distribution of $\hat{\lambda}_0$ is scaled Cauchy

Simulation studies verify that demodulated likelihood yields estimators that seem to have better small sample performance than the classic Whittle likelihood when λ_0 is unknown.

Sampling distribution of estimator

Simulation study: 1000 reps., $N = 4096$, $\lambda_0 = 1/7$, $\delta = 0.4$



Small sample properties

A simulation study: $\lambda_0 = 0.15$, $\delta = 0.1, 0.2, 0.3, 0.4$, $N = 406$, 1000 replicate data sets.

	Adjusted Whittle		Classic Whittle	
δ	Mean	SD	Mean	SD
0.1	0.104	0.0294	0.100	0.0283
0.2	0.199	0.0312	0.194	0.0316
0.3	0.298	0.0313	0.297	0.0325
0.4	0.399	0.0291	0.405	0.0330

Non likelihood estimation

- Geweke-Porter-Hudak (GPH); semi-parametric, generalized least-squares using periodogram near singularity.
- Giraitis-Hidalgo-Robinson; minimize the discrepancy measure

$$G(\delta, \lambda_0) = \frac{1}{M} \sum_{j=0}^{M} \frac{I(j/N)}{S_{GHR}(j/N; \lambda_0, \delta)}.$$

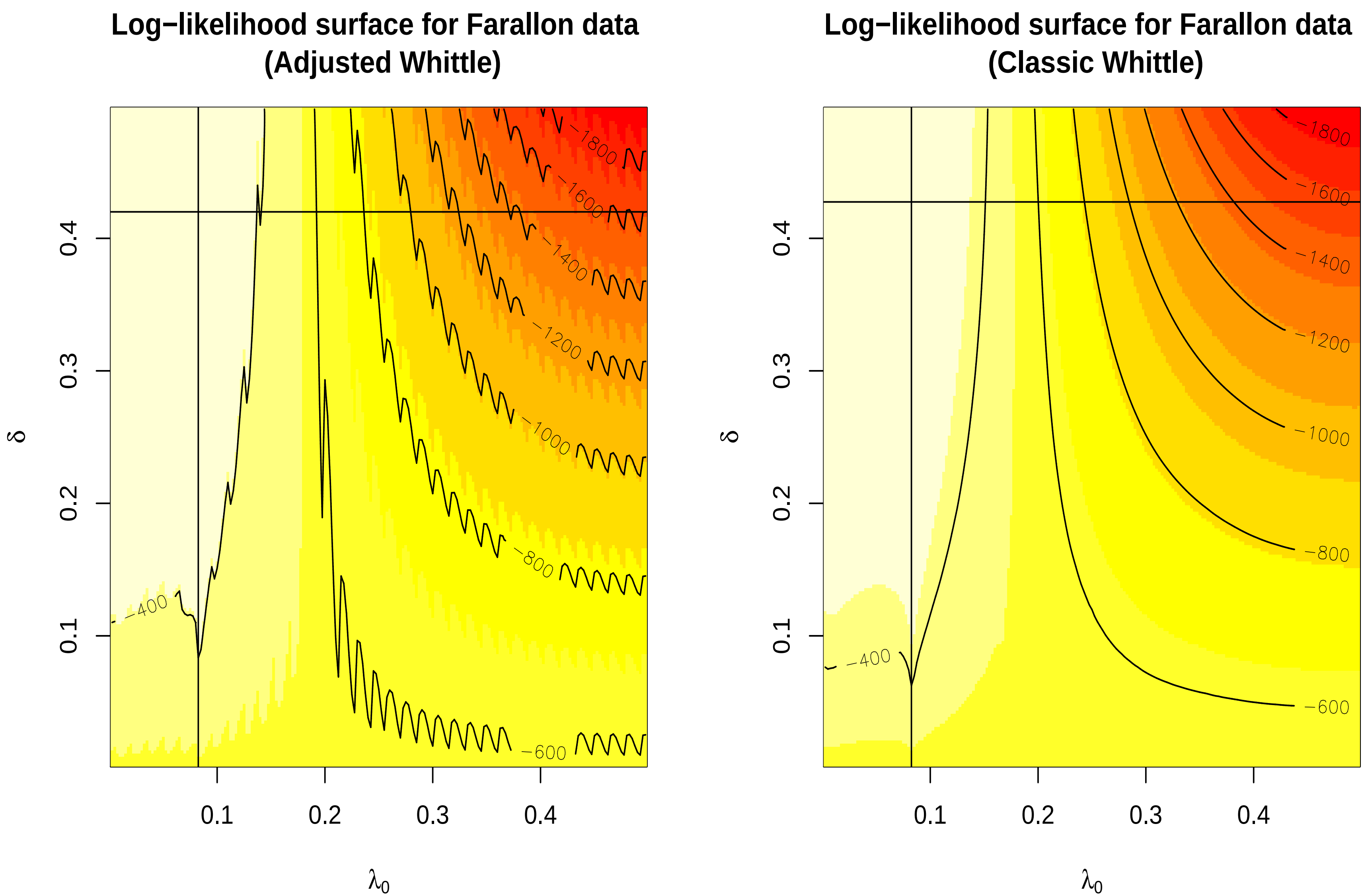
  over a "fine grid" of values for λ_0.
- also common to choose $\hat{\lambda}_0$ equal to the Fourier frequency at which the periodogram achieves its maximum value (Hidalgo-Soulier).

Simulation Study (bias $\times 10^4$).

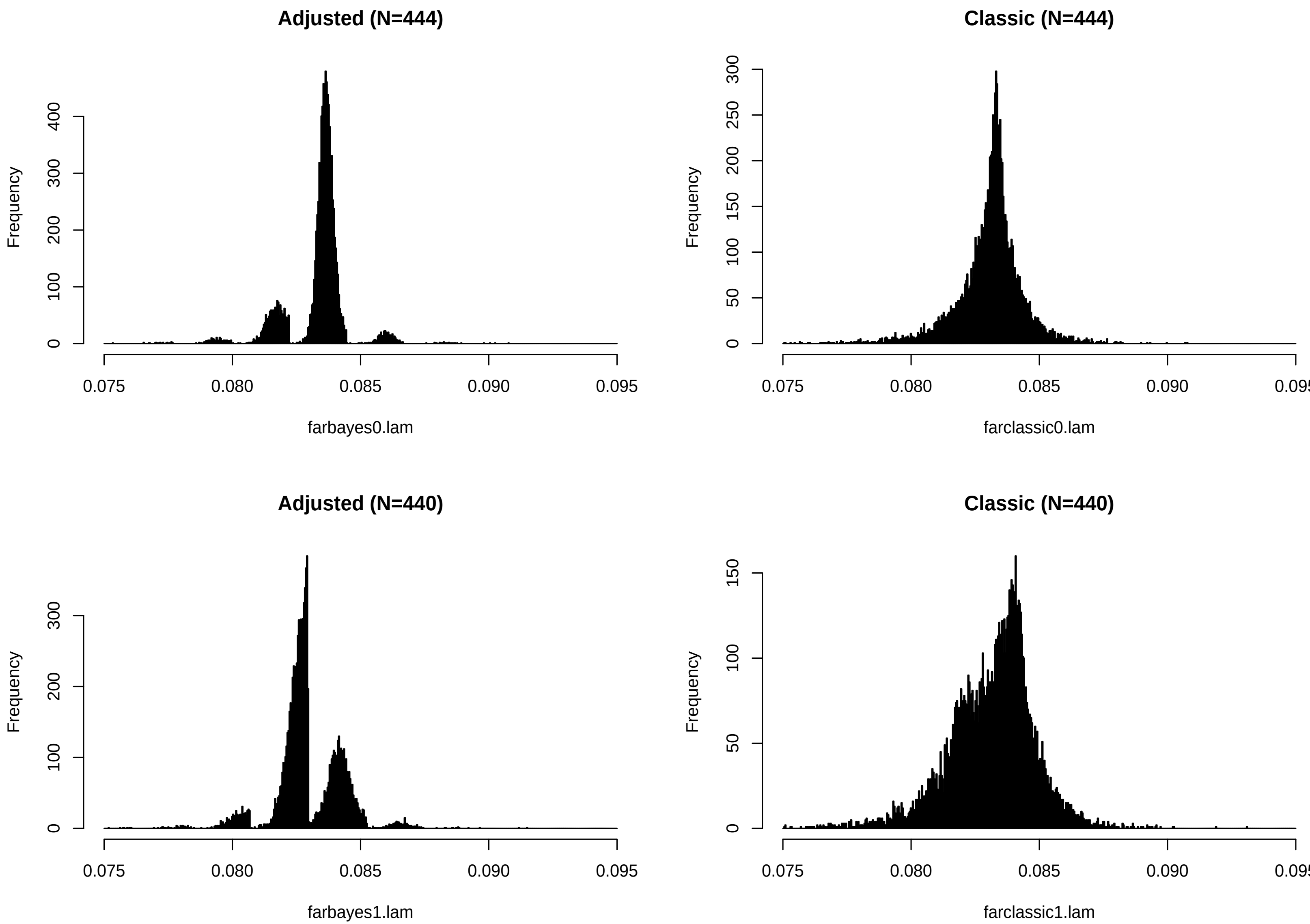
$N = 512$, 500 replications, comparison with other methods.

True Value	ML		GL		HS		PHS	
	Bias	s.d.	Bias	s.d.	Bias	s.d.	Bias	s.d.
$\delta = 0.1$	4	258	-203	289	-583	754	-266	773
λ_0=0.1415	0	54	-4	54	125	869	129	868
$\delta = 0.2$	-8	280	-147	309	-330	761	-64	687
λ_0=0.1415	1	35	-1	37	-26	200	-28	197
$\delta = 0.3$	-25	271	-86	308	-97	677	67	602
λ_0=0.1415	-1	23	-2	25	-3	50	-7	48
$\delta = 0.4$	-18	250	96	348	165	724	137	686
λ_0=0.1415	1	12	0	15	1	23	-5	26
$\delta = 0.45$	-46	223	206	305	544	903	337	772
λ_0=0.1415	0	7	-3	11	-2	16	-3	17

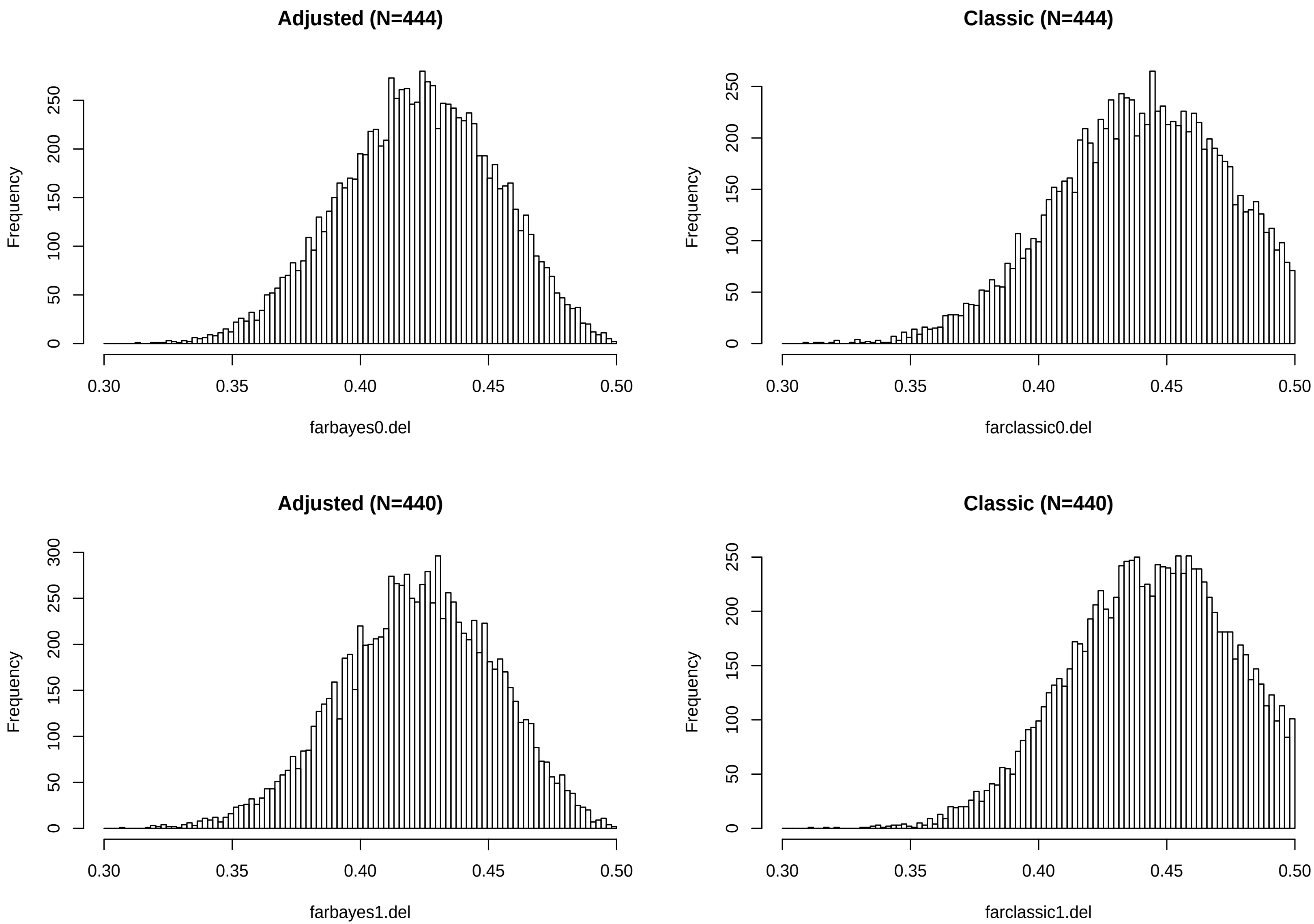
Examples revisited: Farallon Data



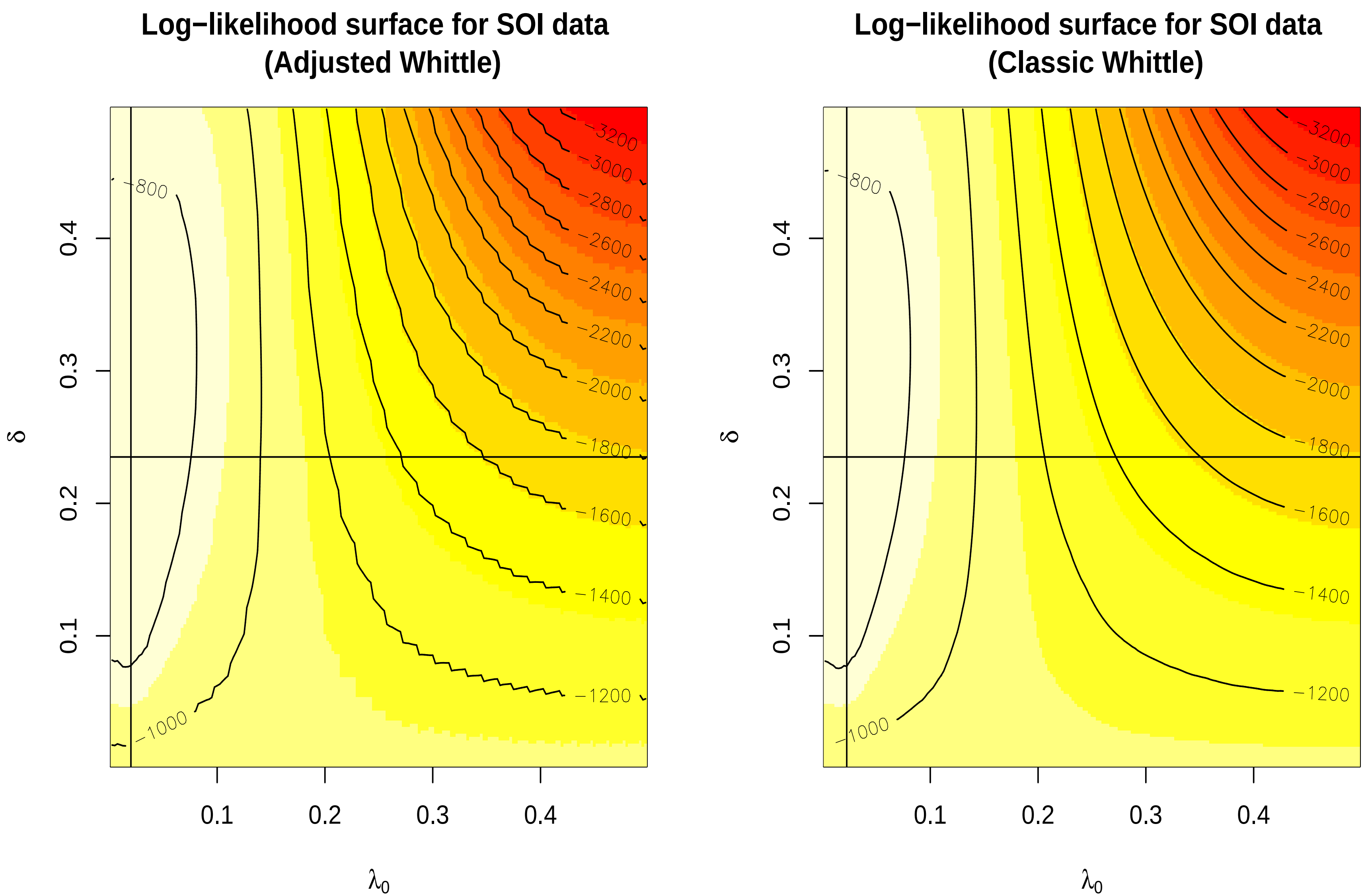
Posterior Distributions : λ_0



Posterior Distributions : δ



Examples revisited: SOI data ($N = 1688$)



GARMA Models

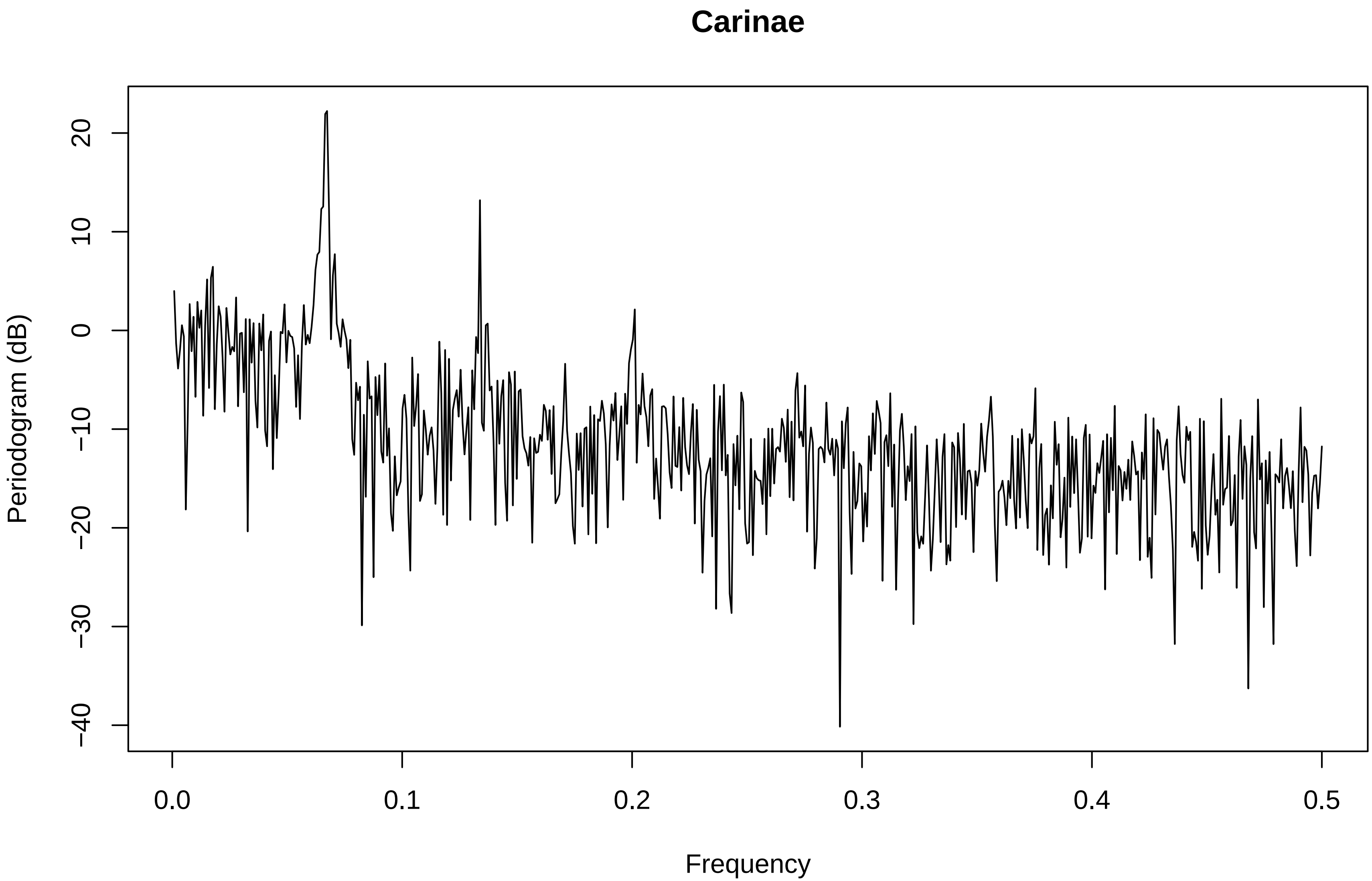
The spectrum of a k-factor GARMA model;

$$S(f) = \sigma_\epsilon^2 \frac{\prod_{j=1}^{q} \left|\left(1 - \zeta_{2j} e^{i2\pi f}\right)\right|^2}{\prod_{j=1}^{p} \left|\left(1 - \zeta_{1j} e^{i2\pi f}\right)\right|^2} \frac{1}{\prod_{j=1}^{k} \left[4\left\{\cos(2\pi f) - \psi_j\right\}^2\right]^{\delta_j}},$$

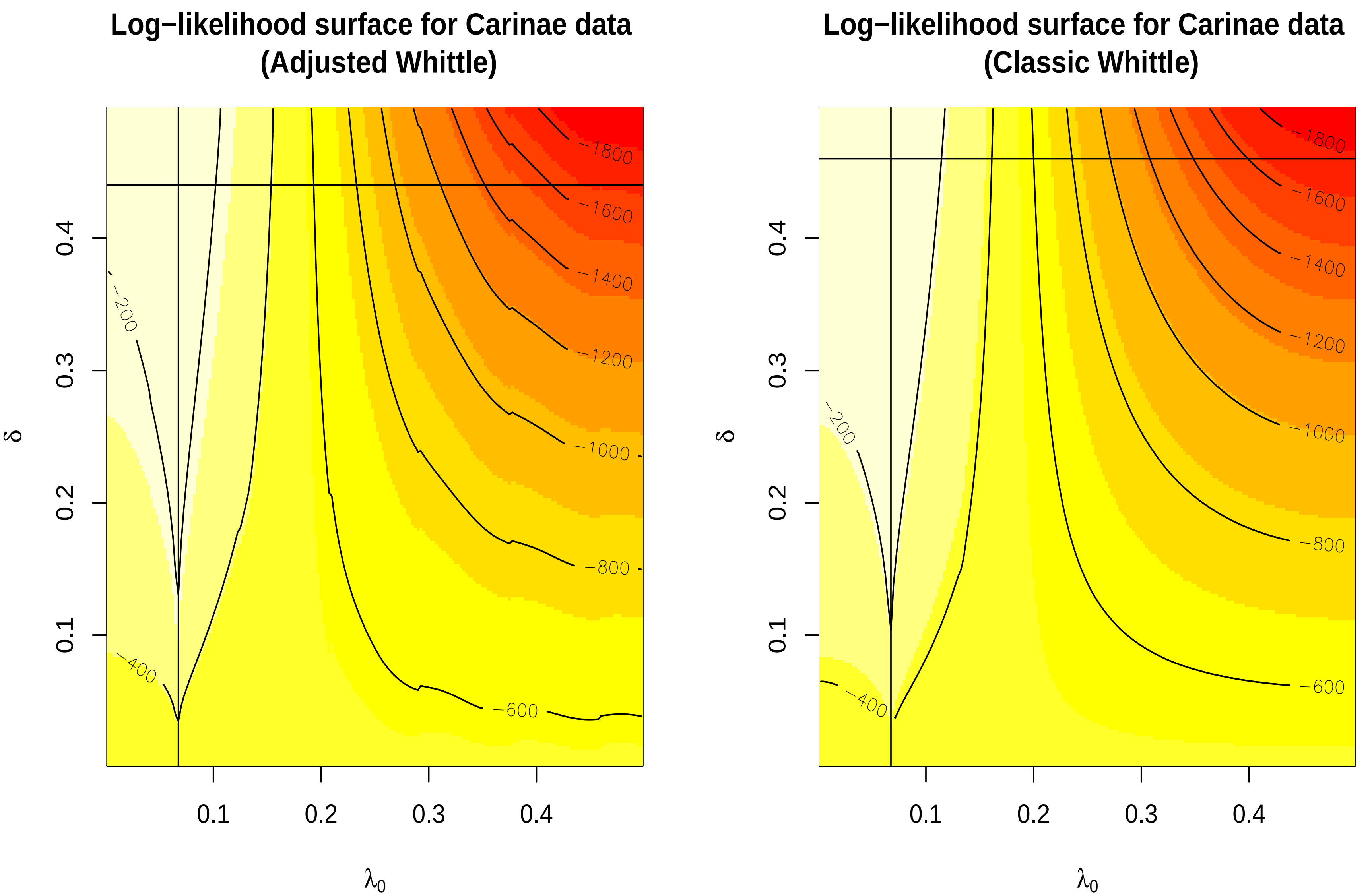
where $\psi_j = \cos(2\pi\lambda_{0j})$ parameterizes the location of the jth singularity in the spectrum.

Use variable dimension MCMC to carry out inference and prediction.

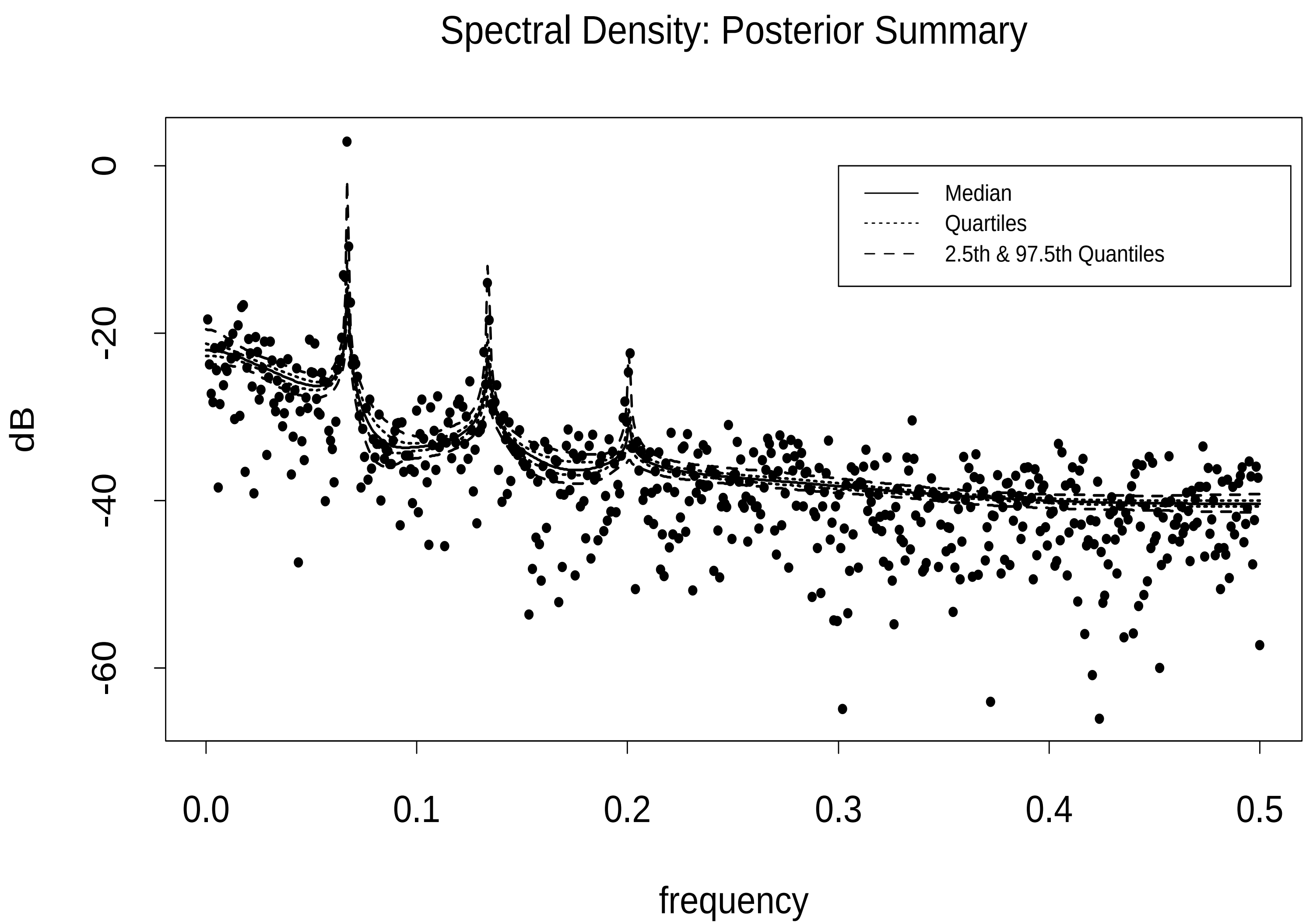
Examples revisited: Carinae data



Carinae data: Pure Gegenbauer



GARMA fit using MCMC



Extensions

- Whittle adjustments for multiple Gegenbauer models
- modelling harmonics
- adjustments to current semiparametric methods
- adjusted continuous Whittle likelihood
- MCMC inference, model selection, imputation of missing values, prediction
- prediction (forecasting) required in time domain
  - inference in frequency domain ?
  - achieved using data augmentation technique in MCMC.