# Parametric model fitting: autoregressive processes

Here we concentrate on zero-mean models of the form

$$X_t - \phi_{1,p} X_{t-1} - \ldots - \phi_{p,p} X_{t-p} = \epsilon_t.$$

As we have seen the corresponding sdf is

$$S(f) = \frac{\sigma_\epsilon^2}{|1 - \phi_{1,p} e^{-i2\pi f} - \ldots - \phi_{p,p} e^{-i2\pi f p}|^2}.$$

This class of models is appealing to use for time series analysis for several reasons:

[1] Any time series with a purely continuous sdf can be approximated well by an AR($p$) model if $p$ is large enough.

[2] There exist efficient algorithms for fitting AR($p$) models to time series.

[3] Quite a few physical phenomena are reverberant and hence an AR model is naturally appropriate.

# A method for estimating the $\{\phi_{j,p}\}$ – Yule-Walker

We start by multiplying the defining equation by $X_{t-k}$:

$$X_t X_{t-k} = \sum_{j=1}^{p} \phi_{j,p} X_{t-j} X_{t-k} + \epsilon_t X_{t-k}.$$

Taking expectations, for $k > 0$:

$$s_k = \sum_{j=1}^{p} \phi_{j,p} s_{k-j}.$$

Let $k = 1, 2, \ldots, p$ and recall that $s_{-\tau} = s_\tau$ to obtain

$$\begin{aligned}
s_1 &= \phi_{1,p} s_0 + \phi_{2,p} s_1 + \ldots + \phi_{p,p} s_{p-1} \\
s_2 &= \phi_{1,p} s_1 + \phi_{2,p} s_0 + \ldots + \phi_{p,p} s_{p-2} \\
\vdots & \quad \vdots \\
s_p &= \phi_{1,p} s_{p-1} + \phi_{2,p} s_{p-2} + \ldots + \phi_{p,p} s_0
\end{aligned}$$

# Yule-Walker

... or in matrix notation,

$$\boldsymbol{\gamma}_p = \Gamma_p \boldsymbol{\phi}_p,$$

where $\boldsymbol{\gamma}_p = [s_1, s_2, \ldots, s_p]^T$; $\boldsymbol{\phi}_p = [\phi_{1,p}, \phi_{2,p}, \ldots, \phi_{p,p}]^T$ and

$$\Gamma_p = \begin{bmatrix} s_0 & s_1 & \cdots & s_{p-1} \\ s_1 & s_0 & \cdots & s_{p-2} \\ \vdots & \vdots & & \vdots \\ s_{p-1} & s_{p-2} & \cdots & s_0 \end{bmatrix}$$

Note: this is a symmetric Toeplitz matrix which we have met already. All elements on a diagonal are the same.

# Yule-Walker

Suppose we don't know the $\{s_\tau\}$, but the mean is zero, then take

$$\hat{s}_\tau = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|},$$

and substitute these for the $s_\tau$'s in $\gamma_p$ and $\Gamma_p$ to obtain $\hat{\gamma}_p, \hat{\Gamma}_p$, from which we estimate $\phi_p$ as $\hat{\phi}_p$:

$$\hat{\phi}_p = \hat{\Gamma}^{-1} \hat{\gamma}_p.$$

# Yule-Walker

Finally, we need to estimate $\sigma_\epsilon^2$. To do so, we multiply the defining equation by $X_t$ and take expectations to obtain

$$\begin{aligned} s_0 &= \sum_{j=1}^{p} \phi_{j,p} s_j + \mathsf{E}\{\epsilon_t X_t\} \\ &= \sum_{j=1}^{p} \phi_{j,p} s_j + \sigma_\epsilon^2, \end{aligned}$$

so that as an estimator for $\sigma_\epsilon^2$ we take

$$\hat{\sigma}_\epsilon^2 = \hat{s}_o - \sum_{j=1}^{p} \hat{\phi}_{j,p} \hat{s}_j.$$

The estimators $\hat{\boldsymbol{\phi}}_p$ and $\hat{\sigma}_\epsilon^2$ are called the Yule-Walker estimators of the AR($p$) parameters.

# Yule-Walker

The estimate of the sdf resulting is

$$\hat{S}(f) = \frac{\hat{\sigma}_\epsilon^2}{\left|1 - \sum_{j=1}^{p} \hat{\phi}_{j,p} e^{-i2\pi f j}\right|^2}.$$

There are two important modifications which we can make to this approach:

[1] We could use for $\{\hat{s}_\tau\}$ a modified autocovariance incorporating tapering:

$$\hat{s}_\tau = \sum_{t=1}^{N-|\tau|} h_t X_t h_{t+|\tau|} X_{t+|\tau|}.$$

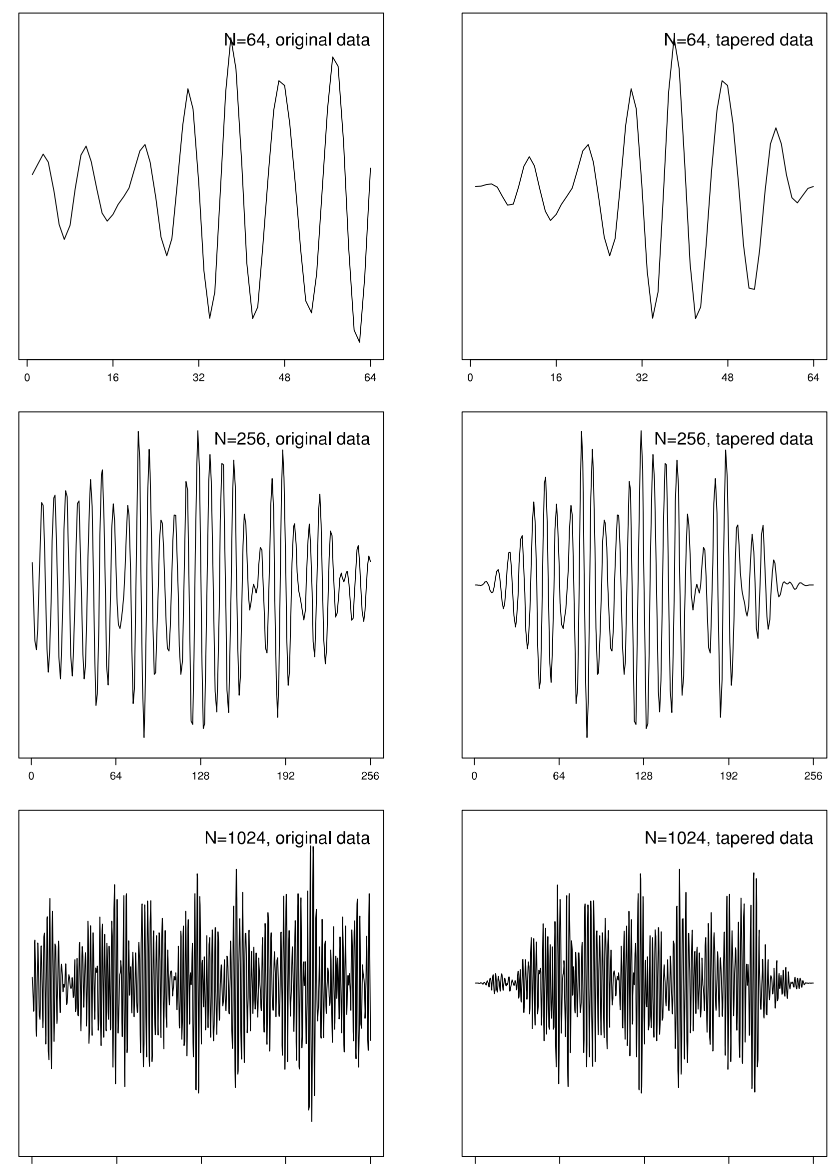
[2] To invert $\hat{\Gamma}_p$ by brute force matrix inversion requires $O(p^3)$ operations. Fortunately, there is an algorithm due to Levinson and Durbin which takes advantage of the highly structured nature of the Toeplitz matrix, and carries out the estimation in $O(p^2)$ or fewer operations.

# Examples:

The AR(4) process again.

- Figure 32: Shows simulations from the AR(4) process defined by,

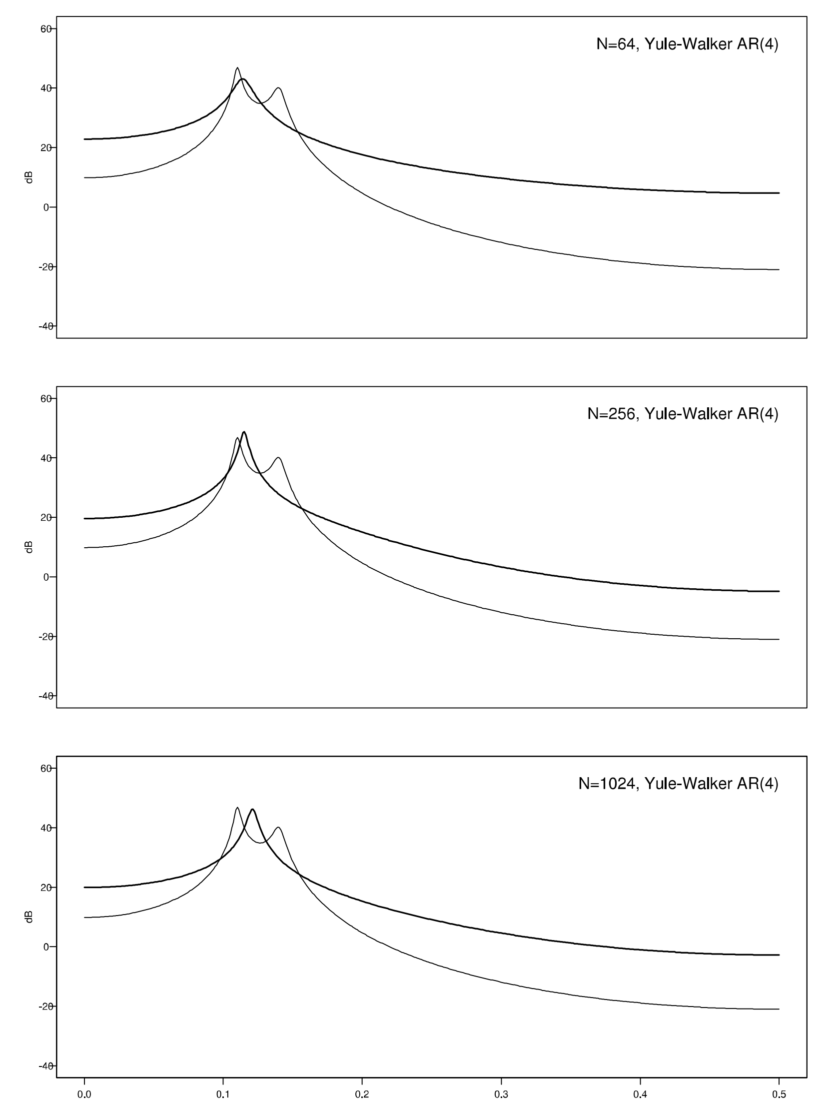
$$X_t = 2.7607X_{t-1} - 3.8106X_{t-2} + 2.6535X_{t-3} - 0.9258X_{t-3} + \epsilon_t$$



- Figure 33: Shows AR(4) processes fitted to the AR(4) data using Yule-Walker method and

$$\hat{s}_\tau = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|}.$$

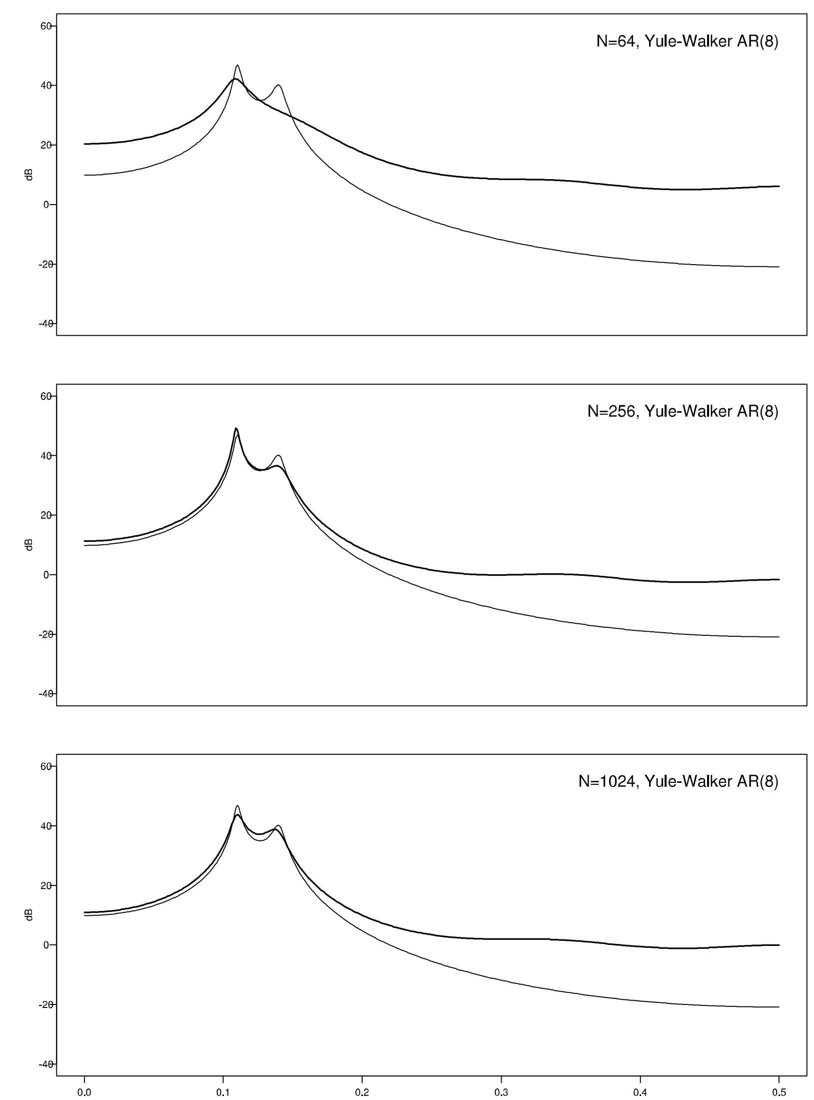
Very poor, even for $N = 1024$.



- Figure 34: Shows AR(8) processes fitted to the AR(4) data using Yule-Walker method and

$$\hat{s}_\tau = \frac{1}{N} \sum_{t=1}^{N-|\tau|} X_t X_{t+|\tau|}.$$

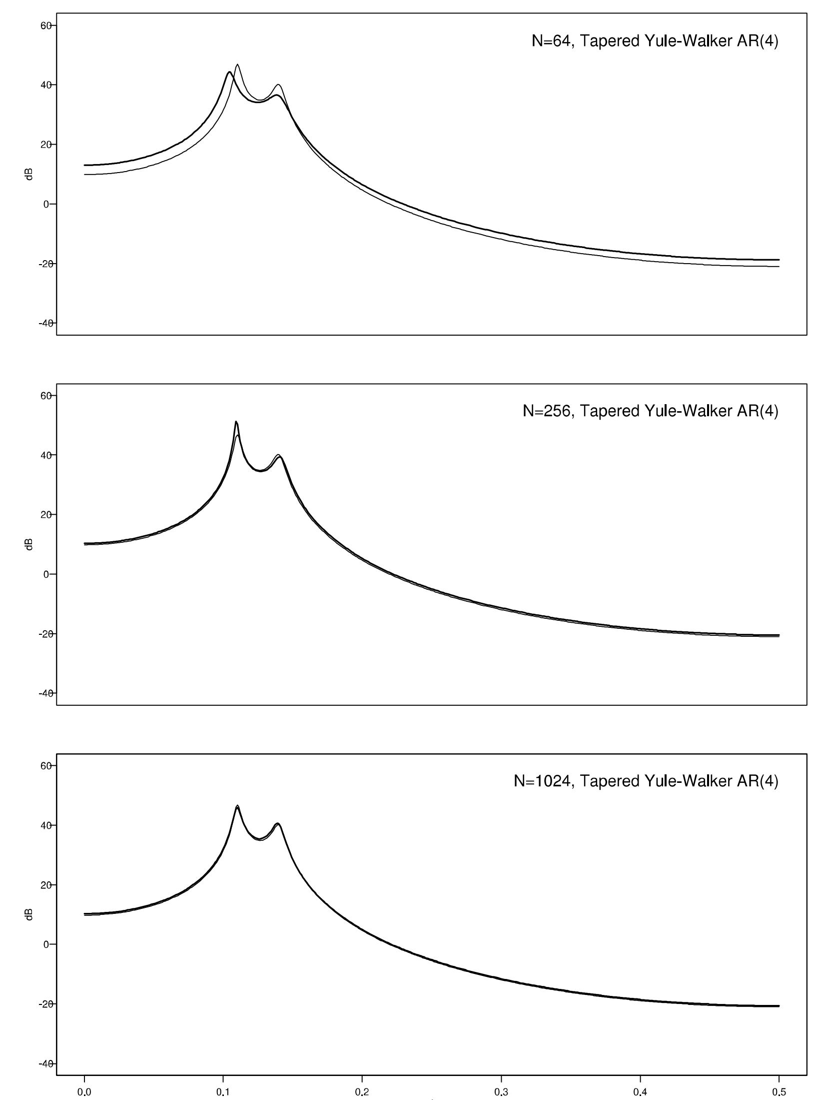
Although the process fitted is not the correct one, the extra parameters have improved the fit.



- Figure 35: Shows AR(4) process fitted to the AR(4) data, using Yule-Walker, but with the 50% split cosine bell taper used:

$$\hat{s}_\tau = \sum_{t=1}^{N-|\tau|} h_t X_t h_{t+|\tau|} X_{t+|\tau|}.$$

The improvement over the other Yule-Walker estimates is dramatic.



The parameter estimates for the fitted AR(4) models when N=1024 are:

| | true | Yule-Walker | tapered Y-W |
|---|---|---|---|
| $\phi_{1,4}$ | 2.7607 | 1.8459 | 2.7636 |
| $\phi_{2,4}$ | -3.8106 | -1.7138 | -3.8108 |
| $\phi_{3,4}$ | 2.6535 | 0.6200 | 2.6502 |
| $\phi_{4,4}$ | -0.9258 | -0.1437 | -0.9211 |
| $\sigma_\epsilon^2$ | 1.0 | 14.9758 | 1.0841 |

# Least squares estimation of the $\{\phi_{j,p}\}$

Let $\{X_t\}$ be a zero-mean AR($p$) process, i.e.,

$$X_t - \phi_{1,p}X_{t-1} - \phi_{2,p}X_{t-2} + \ldots - \phi_{p,p}X_{t-p} = \epsilon_t.$$

We can formulate an appropriate least squares model in terms of data $X_1, X_2, \ldots, X_N$ as follows:

$$\boldsymbol{X}_F = F\boldsymbol{\phi} + \boldsymbol{\epsilon}_F,$$

where,

$$F = \begin{bmatrix} X_p & X_{p-1} & \ldots & X_1 \\ X_{p+1} & X_p & \ldots & X_2 \\ \vdots & & & \vdots \\ X_{N-1} & X_{N-2} & \ldots & X_{N-p} \end{bmatrix}$$

and,

$$\boldsymbol{X}_F = \begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_N \end{bmatrix}; \quad \boldsymbol{\phi} = \begin{bmatrix} \phi_{1,p} \\ \phi_{2,p} \\ \vdots \\ \phi_{p,p} \end{bmatrix}; \quad \boldsymbol{\epsilon}_F = \begin{bmatrix} \epsilon_{p+1} \\ \epsilon_{p+2} \\ \vdots \\ \epsilon_N \end{bmatrix}.$$

We can thus estimate $\phi$ by finding that $\phi$ such that

$$\begin{aligned} \mathrm{SS}_F(\phi) &= \sum_{t=p+1}^{N} \left( X_t - \sum_{k=1}^{p} \phi_{k,p} X_{t-k} \right)^2 \qquad \left[ = \sum_{t=p+1}^{N} \epsilon_t^2 \right] \\ &= (\boldsymbol{X}_F - F\phi)^T (\boldsymbol{X}_F - F\phi) \end{aligned}$$

is minimized.

If we denote the vector that minimizes the above as $\hat{\phi}_F$, standard least squares theory tells us that it is given by

$$\hat{\phi}_F = (F^T F)^{-1} F^T \boldsymbol{X}_F.$$

Note: convince yourselves of this using the fact that:

$$\frac{\partial}{\partial \boldsymbol{x}} (A\boldsymbol{x} + \boldsymbol{b})^T (A\boldsymbol{x} + \boldsymbol{b}) = 2A^T (A\boldsymbol{x} + \boldsymbol{b}).$$

We can estimate the innovations variance $\sigma_\epsilon^2$ by the usual estimator of the residual variation, namely

$$\hat{\sigma}_F^2 = \frac{(\boldsymbol{X}_F - F\hat{\phi}_F)^T (\boldsymbol{X}_F - F\hat{\phi}_F)}{(N - 2p)}.$$

(Note: there are $N - p$ effective observations, and $p$ parameters are estimated).

The estimator $\hat{\phi}_F$ is known as the forward least squares estimator of $\phi$.

But a stationary Gaussian AR($p$) process also has a "time reversed" formulation, so we could rewrite the least squares problem as

$$\boldsymbol{X}_B = B\phi + \boldsymbol{\epsilon}_B,$$

where,

$$B = \begin{bmatrix} X_2 & X_3 & \cdots & X_{p+1} \\ X_3 & X_4 & \cdots & X_{p+2} \\ \vdots & & & \vdots \\ X_{N-p+1} & X_{N-p+2} & \cdots & X_N \end{bmatrix}$$

and,

$$\boldsymbol{X}_B = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_{N-p} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\epsilon}_B = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{N-p} \end{bmatrix}.$$

The function of $\phi$ to be minimized is now

$$\begin{aligned} \mathrm{SS}_B(\phi) &= \sum_{t=1}^{N-p} \left( X_t - \sum_{k=1}^{p} \phi_{k,p} X_{t+k} \right)^2 \\ &= (\boldsymbol{X}_B - B\phi)^T (\boldsymbol{X}_B - B\phi) \end{aligned}$$

The backward least squares estimator of $\phi$ is then given by

$$\hat{\phi}_B = (B^T B)^{-1} B^T \boldsymbol{X}_B.$$

The corresponding estimator of the innovations variance is

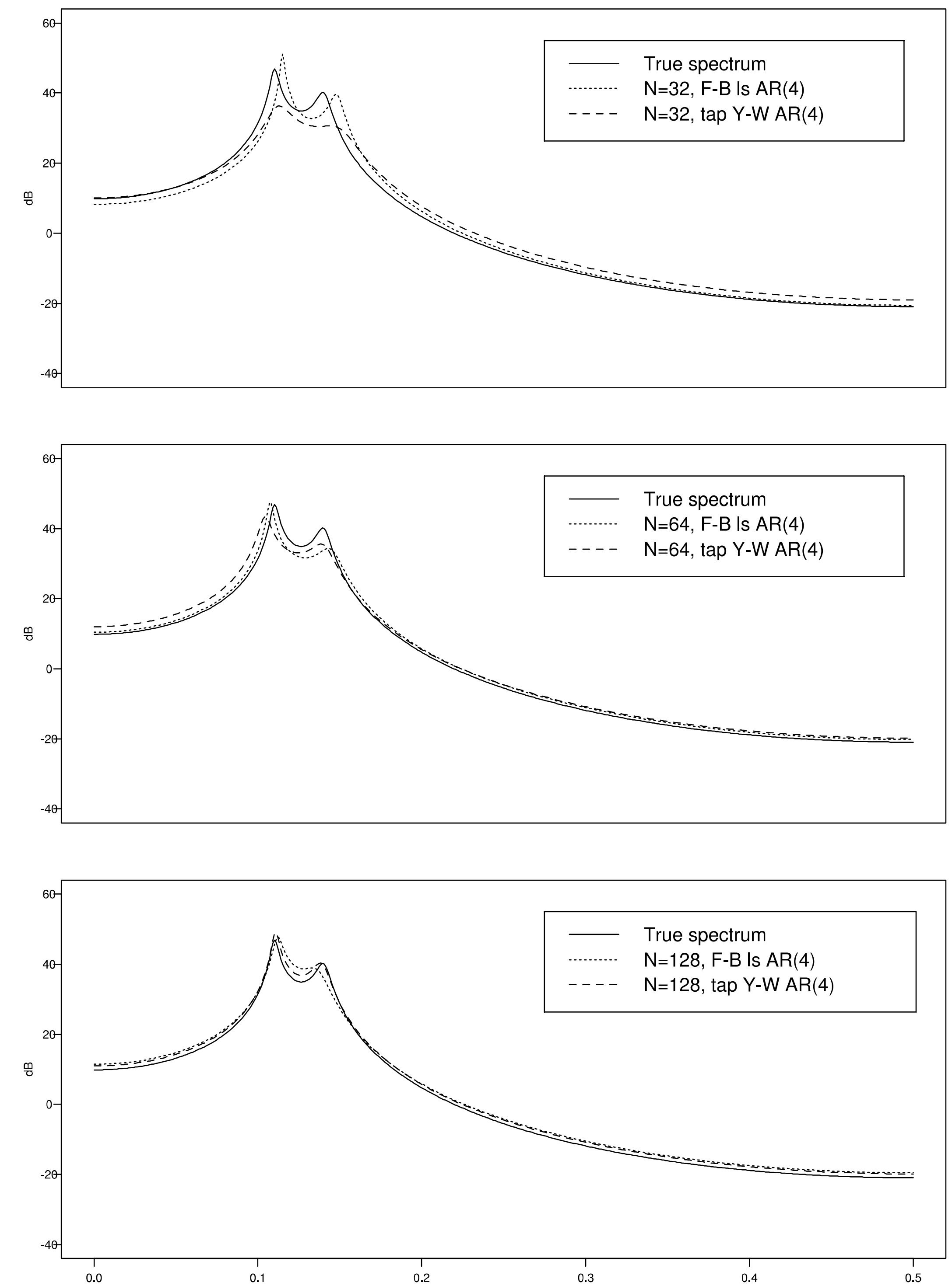
$$\hat{\sigma}_B^2 = \frac{(\boldsymbol{X}_B - B\phi)^T (\boldsymbol{X}_B - B\phi)}{(N - 2p)}.$$

The vector $\hat{\phi}_{FB}$ that minimizes

$$SS_F(\phi) + SS_B(\phi)$$

is called the forward/backward least squares estimator, and Monte-Carlo studies indicate that it performs better than forward or backward least squares.

Figure 36 shows the AR(4) spectra corresponding to the forward/backward least squares estimates of $\phi$ and tapered Yule-Walker estimates for comparison.



# NOTES

[1] $\hat{\phi}_{FB}, \hat{\phi}_B$ and $\hat{\phi}_F$ produce estimated models which need not be stationary. This may be a concern for prediction, however, for spectral estimation, the parameter values will still produce a valid sdf (i.e., nonnegative everywhere, symmetric about the origin and integrates to a finite number).

[2] The Yule-Walker estimates can be formulated as a least squares problem. Consider adding zeros to our observations $X_1, X_2, \ldots, X_N$, both at the beginning and end of the data, to give:

$$\boldsymbol{X}_{YW} = W\phi + \boldsymbol{\epsilon}_{YW},$$

where,

$$W = \begin{bmatrix} 0 & 0 & 0 & \ldots & \ldots & 0 \\ X_1 & 0 & 0 & \ldots & \ldots & 0 \\ X_2 & X_1 & 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & & & & \vdots \\ X_{p-1} & \vdots & & & & 0 \\ X_p & X_{p-1} & \ldots & \ldots & \ldots & X_1 \\ \vdots & \vdots & & & & \vdots \\ X_N & X_{N-1} & \ldots & \ldots & \ldots & X_{N-p+1} \\ 0 & X_N & & & & X_{N-p+2} \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & & & & X_N \end{bmatrix}$$

and,

$$\boldsymbol{X}_{YW} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \text{and} \quad \boldsymbol{\epsilon}_{YW} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Note that,

$$\frac{1}{N} W^T W = \begin{bmatrix} \hat{s}_0^{(p)} & \hat{s}_1^{(p)} & \dots & \hat{s}_{p-1}^{(p)} \\ \hat{s}_1^{(p)} & \ddots & & \\ \vdots & \ddots & \ddots & \\ \hat{s}_{p-1}^{(p)} & \dots & \dots & \hat{s}_0^{(p)} \end{bmatrix} = \hat{\Gamma}_p$$

and

$$\frac{1}{N} W^T \boldsymbol{X}_{YW} = \begin{bmatrix} \hat{s}_1^{(p)} \\ \vdots \\ \hat{s}_p^{(p)} \end{bmatrix} = \hat{\gamma}_p,$$

so that

$$(W^T W)^{-1} W^T \boldsymbol{X}_{YW} = (\hat{\Gamma}_p)^{-1} \hat{\gamma}_p.$$

which is identical to the Yule-Walker estimate.