

## Expedient #2: Ignore the off-diagonals

I'll develop this to help explain the DLM puzzle, which used the model

$$y_t = s_t + h_t + r_t + m_t + v_t$$

with the components modeled as follows:

- ▶ signal  $s_t$ , local linear model
- ▶ heartbeat  $h_t$ , quasi-cyclic, nominal period 2.78
- ▶ respiration  $r_t$ , quasi-cyclic, nominal period 18.75
- ▶ mystery  $h_t$ , quasi-cyclic, nominal period 117 (Model 2 only)
- ▶ iid error  $v_t$

and  $t = 0, \dots, T = 649$

## MLM representation: Signal

This component's FE and RE design matrices are

$$\mathbf{X}_s = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ \vdots & \vdots \\ 1 & T \end{pmatrix}, \quad \mathbf{Z}_s = \begin{pmatrix} 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 2 & 1 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & & \ddots & \vdots \\ T-1 & T-2 & T-3 & T-4 & \dots & 1 \end{pmatrix}$$

This is a penalized spline with a truncated-line basis and a knot at every observation.

## MLM representation: Quasi-cyclic component

$\delta = 2\pi/\text{period}$ ; the FEs have the component's nominal frequency

$$\begin{bmatrix}
 1 & 0 \\
 \cos \delta & \sin \delta \\
 \cos 2\delta & \sin 2\delta \\
 \cos 3\delta & \sin 3\delta \\
 \vdots & \\
 \cos T\delta & \sin T\delta
 \end{bmatrix}
 \begin{array}{l}
 \leftarrow \text{FE design matrix } \mathbf{X}_k \\
 \downarrow \text{RE design matrix } \mathbf{Z}_k
 \end{array}$$
  

$$\begin{bmatrix}
 0 & 0 & 0 & 0 & 0 & 0 \\
 1 & 0 & 0 & 0 & \dots & 0 \\
 \cos \delta & \sin \delta & 1 & 0 & \dots & 0 \\
 \cos 2\delta & \sin 2\delta & \cos \delta & \sin \delta & \dots & 0 \\
 \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
 \cos(T-1)\delta & \sin(T-1)\delta & \cos(T-2)\delta & \sin(T-2)\delta & \dots & 1 & 0
 \end{bmatrix}$$

# The model and exact RL

Model 2 is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{s,m,r,h} \mathbf{Z}_k \mathbf{u}_k + \boldsymbol{\epsilon},$$

- ▶  $\mathbf{X} = (\mathbf{X}_s | \mathbf{X}_m | \mathbf{X}_r | \mathbf{X}_h)$  is  $650 \times 8$
- ▶  $\mathbf{Z}_s$  is  $650 \times 648$ ,  $\mathbf{u}_s$  is  $648 \times 1$ ;  $\text{var}(u_{sj}) = \sigma_{ss}^2$
- ▶  $\mathbf{Z}_k, k = m, r, h$  is  $650 \times 1298$ ,  $\mathbf{u}_k$  is  $1298 \times 1$ ;  $\text{var}(u_{kj}) = \sigma_{sk}^2$
- ▶  $\mathbf{R} = \sigma_e^2 \mathbf{I}_{650}$

Define  $\mathbf{K}$   $650 \times 642$  with orthonormal columns  $\ni \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{K}\mathbf{K}'$ .

The RL is the likelihood arising from  $\mathbf{K}'\mathbf{y}$ .

(Model 1 omits mystery;  $\mathbf{X}_m$  and  $\mathbf{Z}_m$  are omitted and  $\mathbf{K}$  changes a bit.)

## The log RL and the idea of Expedient #2

The log RL is the log likelihood arising from  $\mathbf{K}'\mathbf{y}$ :

$$-0.5 \log \left| \sum_k \sigma_{sk}^2 \mathbf{K}'\mathbf{z}_k \mathbf{z}_k' \mathbf{K} + \sigma_e^2 \mathbf{I}_{642} \right|$$
$$-0.5 \mathbf{y}' \mathbf{K} \left[ \sum_k \sigma_{sk}^2 \mathbf{K}'\mathbf{z}_k \mathbf{z}_k' \mathbf{K} + \sigma_e^2 \mathbf{I}_{642} \right]^{-1} \mathbf{K}'\mathbf{y}.$$

We need to make the  $\mathbf{K}'\mathbf{z}_k \mathbf{z}_k' \mathbf{K}$  and  $\mathbf{I}_{642}$  diagonal; it's impossible.

**Idea:**

- ▶ The columns of  $\mathbf{K}'\mathbf{z}_k$  are  $\approx$  sinusoidal with frequencies starting just above the nominal frequency and increasing.
- ▶ Transform  $\mathbf{K}'\mathbf{y}$  to diagonalize  $\mathbf{I}_{642}$  and  $\mathbf{K}'\mathbf{z}_s \mathbf{z}_s' \mathbf{K}$ ; the other 3  $\mathbf{K}'\mathbf{z}_k \mathbf{z}_k' \mathbf{K}$  should have mostly small off-diagonals.
- ▶  $\Rightarrow$  approximate the RL by ignoring those off-diagonals.
- ▶ The components have different nominal frequencies, so they should have large diagonals for different transformed observations.

## The approximate RL, with the desired simple form

Let  $\mathbf{K}'\mathbf{Z}_s\mathbf{Z}_s'\mathbf{K} = \mathbf{\Gamma}\mathbf{A}_s\mathbf{\Gamma}'$  (spectral decomposition).

$$\log RL \propto -0.5 \log \left| \sum_k \sigma_{sk}^2 \mathbf{A}_k + \sigma_e^2 \mathbf{I}_{642} \right| \\ -0.5 \mathbf{y}' \mathbf{K} \mathbf{\Gamma} \left[ \sum_k \sigma_{sk}^2 \mathbf{A}_k + \sigma_e^2 \mathbf{I}_{642} \right]^{-1} \mathbf{\Gamma}' \mathbf{K}' \mathbf{y}$$

where  $\mathbf{A}_k = \mathbf{\Gamma}' \mathbf{K}' \mathbf{Z}_k \mathbf{Z}_k' \mathbf{K} \mathbf{\Gamma}$ ,  $k = s, m, r, h$  with diagonal elements  $a_{jk}$ .

Approximate  $\mathbf{A}_m$ ,  $\mathbf{A}_r$ ,  $\mathbf{A}_h$  by their diagonals to give

$$\log ARL = K \quad -0.5 \sum_j \log \left( \sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2 \right) \\ -0.5 \sum_j \hat{v}_j^2 / \left( \sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2 \right),$$

For “data”  $\hat{\mathbf{v}} = \mathbf{\Gamma}' \mathbf{K}' \mathbf{y}$ , this is the likelihood from a GLM with gamma errors, identity link, and  $E(\hat{v}_j^2) = \sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2$ .

# The approximate RL is the RL for an approximate model

In effect, we're replacing the DLM

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \sum_{s,m,r,h} \mathbf{z}_k \mathbf{u}_k + \boldsymbol{\epsilon}, \quad \text{var}(u_{kj}) = \sigma_{sk}^2$$

with this approximate model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{K}\boldsymbol{\Gamma}\mathbf{v} + \boldsymbol{\epsilon}$$

where

- $\mathbf{K}$   $n \times (n - 8)$  has orthonormal columns;  $\mathbf{K}\mathbf{K}' = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$
- $\boldsymbol{\Gamma}$   $(n - 8) \times (n - 8)$  has columns the eigenvectors of  $\mathbf{K}'\mathbf{Z}_s\mathbf{Z}'_s\mathbf{K}$ .
- $\mathbf{v}$   $(n - 8) \times 1$  has  $\text{cov}(v_j) = \sum_{k=s,m,r,h} a_{jk}\sigma_{sk}^2$

For signal,  $\mathbf{K}\boldsymbol{\Gamma}\text{diag}(a_{js})\boldsymbol{\Gamma}'\mathbf{K}' = \mathbf{K}\mathbf{K}'\mathbf{Z}_s\mathbf{Z}'_s\mathbf{K}\mathbf{K}'$  by construction.

For other components,  $\mathbf{K}\boldsymbol{\Gamma}\text{diag}(a_{jk})\boldsymbol{\Gamma}'\mathbf{K}' \approx \mathbf{K}\mathbf{K}'\mathbf{z}_k\mathbf{z}'_k\mathbf{K}\mathbf{K}' \dots$  we hope.

## Are the (ignored) off-diagonals small? Mostly yes.

Each  $\mathbf{A}_k = \mathbf{\Gamma}'\mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K}\mathbf{\Gamma}$  has 205,761 distinct off-diagonals

Percentiles of absolute off-diagonals of  $\mathbf{A}_k$ :

- ▶ Mystery: 99.9<sup>th</sup> 0.064; 99.99<sup>th</sup> 0.46; max 0.72.
- ▶ Respiration: 95<sup>th</sup> 0.060; 99<sup>th</sup> 0.32; max 0.80.
- ▶ Heartbeat: 50<sup>th</sup> 5e-6, 55<sup>th</sup> 0.11, 75<sup>th</sup> 0.46, 95<sup>th</sup> 0.71, max 0.82.

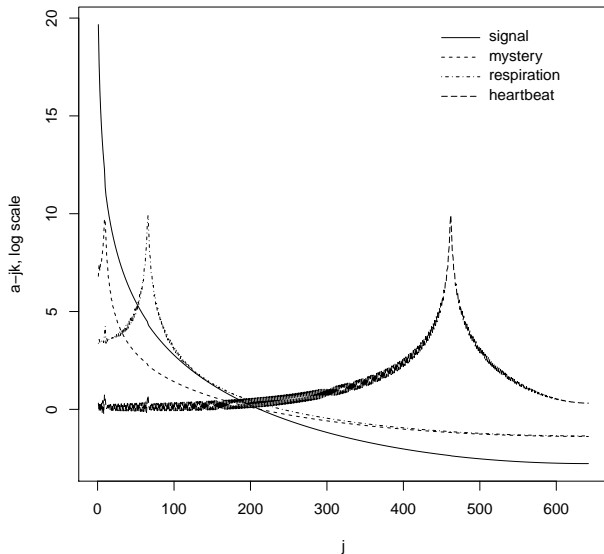
(No surprises. Other models to be discussed have similar off-diagonals.)

Good enough . . . now let's see what happens.

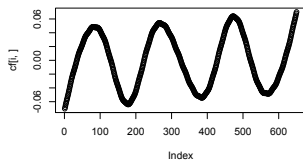
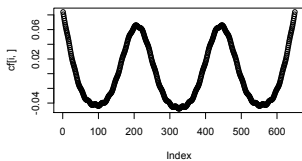
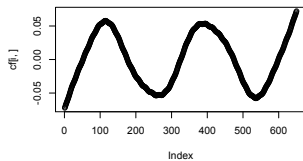
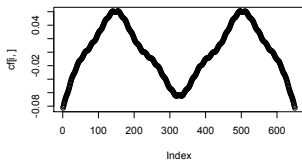
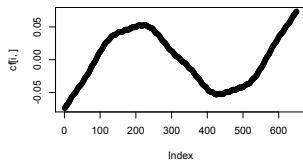
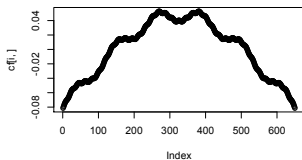


# The canonical predictors and their $a_{kj}$

For Model 2, here are the  $a_{kj}$



# Canonical predictors (columns of $\mathbf{K}\Gamma$ ) with big $a_{js}$



## Canonical predictors with large $a_{jk}$ for other components

Dominant periods or wavelengths (WL) for each component's large  $a_{jk}$

Mystery, nominal WL 117:

- ▶  $j = 7:13$  have WL 168.8 135 135 [117 missing] 96.4 96.4 96.4 84.4

Respiration, nominal WL 18.75:

- ▶  $j = 64:68$  have WL 19.3 19.3 [18.75 missing] 18.2 18.2 17.8

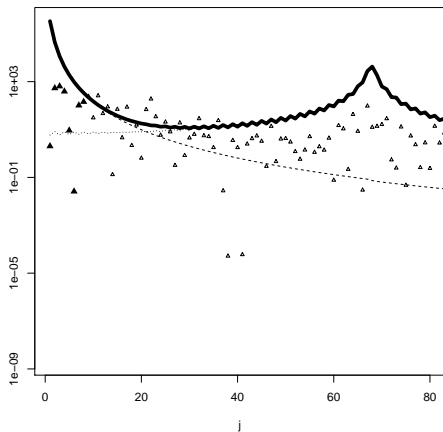
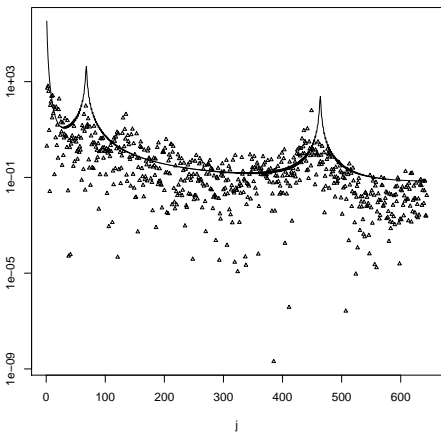
Heartbeat, nominal WL 2.78:

- ▶  $j = 459:463$  have WL 2.80 2.80 2.789 [2.78 missing] 2.77 2.77

Each component's largest  $a_{jk}$  are for  $j$  (canonical predictor,  $\hat{v}_j$ ) with dominant WL close to the component's nominal WL.

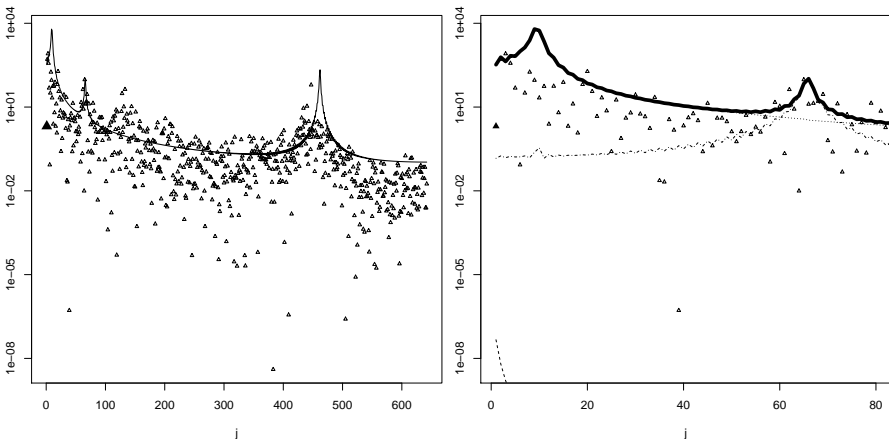
# Fit Model 1, plot $\hat{v}_j^2$ and (approx) model fit to them

Vertical axis:  $\hat{v}_j^2$  (triangles) and fitted values (lines), log scale.



The fit is poor at:  $j$  small;  $j$  near resp peak; resp's two "echo" peaks  
Both sig and resp peaks are pulled up to fit  $\hat{v}_j^2$  for  $j \approx 8-30$ .

## Add mystery (Model 2), plot $\hat{v}_j^2$ & (approx) model fit



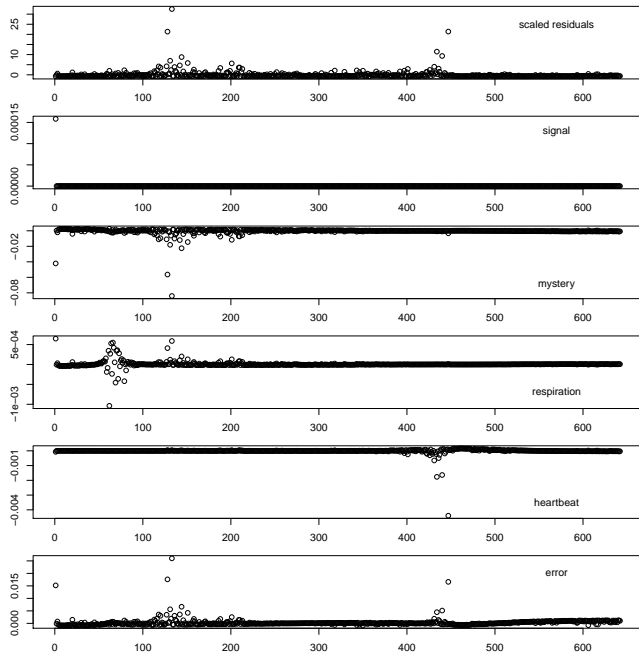
Good: Mystery captures  $j \approx 8-30$ , so respiration fits its first peak better

Bad: Mystery's fit has a peak where the  $\hat{v}_j^2$  don't have a peak;

Respiration's two "echo" peaks are still ignored

Signal is wiped out because  $\hat{v}_1^2$  (big triangle) is so small.

# Scaled residuals (top) and approx leave-one-out changes



## Comments on the preceding slide

$\hat{v}_1^2$  has a big influence: removing it increases signal, respiration, and error, and reduces mystery.

Respiration's first "echo" peak,  $j \approx 135-140$ , is influential: removing those  $j$  reduces mystery and increases respiration and error.

Heartbeat, as usual, is off in its own world.

## Can we make this fit better & estimate signal sensibly?

Hypothesis: Lack of fit to signal and resp forces mystery's  $\sigma_{sm}^2$  upward to "fix" the bad fit, which in turn pushes down  $\sigma_{ss}^2$  and  $\sigma_{sr}^2$ .

Possible solutions:

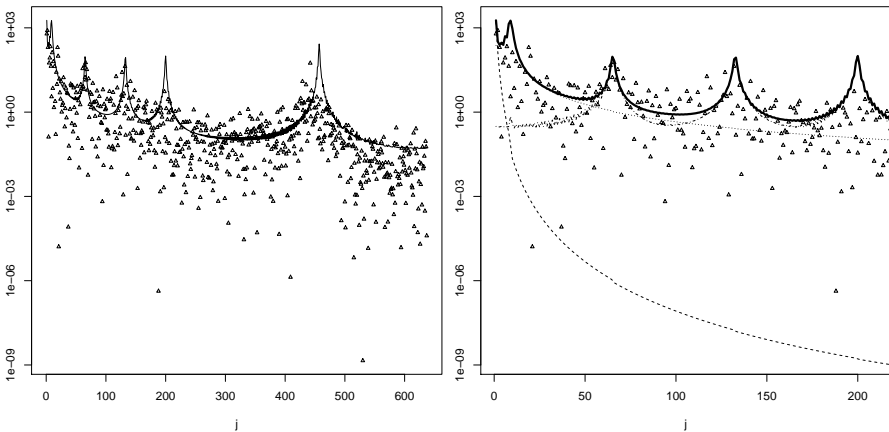
- ▶ Add 2nd and 3rd harmonics for respiration.
- ▶ Make the 1<sup>st</sup> canonical predictor a FE (omit  $\hat{v}_1^2$ ).
  - ▶ We can't do this, because this model is approximate.
  - ▶ We *can* do something that has almost the same effect: replace signal's locally linear model with a locally quadratic model.

Either of these alone, or both together, produce the desired result.

- ▶ Model 1 still has extra junk in the signal fit.
- ▶ But Model 2 now gives a sensible signal fit.
- ▶ Signal's  $\sigma_{ss}^2$  is still sensitive to  $\hat{v}_1^2$ ,  $\hat{v}_2^2$ ; mystery & resp are less so.
- ▶ Mystery's peak still doesn't fit anything, but at least it's smaller.



## Model 2 + quad sig + 3-harm resp: $\hat{v}_j^2$ & (approx) fit



Good: As described.

Bad: Mystery's fit still has a peak where the  $\hat{v}_j^2$  don't have a peak.

The fit's a bit high at respiration's second "echo" peak.

# Stuff I've been sweeping under the rug (1)

Does this make sense as an approach to doing data analysis?

Right now I'm trying to

- ▶ understand why these models fit the way they do
- ▶ devise diagnostics that show lack of fit and suggest models that do fit.

I feel OK about this analysis though I arguably worked to a pre-determined conclusion about signal, which seems reasonable in this case.

## Stuff I've been sweeping under the rug (2)

These models are extremely fussy to fit.

- ▶ First Ellie Duffy and I use Petris's dlm package in R (likelihood).
- ▶ Then we try to maximize the exact RL (ERL).
- ▶ Then we try to maximize the approximate RL (ARL).

We parameterize using log variances

- ▶ This seems necessary because they differ by many orders of mag.
- ▶ BUT you can't get a zero variance estimate this way.
- ▶ The resulting functions are very flat in variances that probably should have zero estimates (usually error, often signal).

For some datasets the ARL appears to have many local maxima; the likelihood and ERL appear to have fewer.

For some models I haven't been able to get a max for the ERL.

## Stuff I've been sweeping under the rug (3)

The analyses I've shown are for Region #1 in Michael Lavine's dataset.

Ellie Duffy has fit the same models to several other regions.

SO FAR, the results are not tidy like the ones I've shown you.

We've been diverted trying to understand how many local maxima the likelihood, ERL, and ARL can have.

Now that

- we've corrected a bug in my leave-one-out code

- we understand (some) how badly behaved these functions are

maybe we can see whether this expedient is useful for the other series.