have contours parallel to the z_1 axis, as in all three panels of Figure 17.2. When these two sets of **u**-free terms and the mixed terms are multiplied together, they produce the two-legged contour plots for Classifications A and B. Classification C's contour plot has only one "leg" parallel to the z_1 axis because it has no **u**-free terms for z_1.

This leads to a secondary puzzle: If the **u**-free terms dominate the marginal posterior of (z_1, z_2), why don't Figures 17.2a,b have "legs" pointing down and to the right, giving substantial posterior mass to small values of z_2 and z_1 respectively? To understand this, we need to examine the marginal posterior itself, equation (17.38). The intuition is that as either z_1 or z_2 takes increasingly large values — as either class of neighbor pairs is shrunk more stringently — the marginal posterior declines slowly and eventually becomes flat. This happens because past a certain point, neighbors have already been shrunk together as much as they practically can be and a further increase in z_1 or z_2 has no practical effect. Mechanically, it happens because $\left[b_e + 0.5(\mathbf{y}'\mathbf{y} - \mathbf{y}'[\mathbf{I}_n + e^{z_1}\mathbf{Q}_1 + e^{z_2}\mathbf{Q}_2]^{-1}\mathbf{y})\right]^{-(n-I)/2-a_e}$ converges to a finite positive constant while $\prod_{j=1}^{n-I}(e^{z_1}d_{1j} + e^{z_2}d_{2j})^{0.5}$ and $|\mathbf{I}_n + e^{z_1}\mathbf{Q}_1 + e^{z_2}\mathbf{Q}_2|^{-0.5}$ offset each other. This flattening can be seen in the upper right corner of Figures 17.2a,b and the upper half of Figure 17.2c. By contrast, as z_1 takes increasingly negative values, the marginal posterior drops off precipitiously if there are **u**-free terms for z_1 and analogously for z_2. Mechanically, this happens because

$$\prod_{j=1}^{n-I}(e^{z_1}d_{1j} + e^{z_2}d_{2j})^{0.5} \propto e^{n_1 z_1/2}e^{n_2 z_2/2} \prod_{\text{mixed terms}} (e^{z_1}d_{1j} + e^{z_2}d_{2j})^{0.5} \qquad (17.41)$$

where n_k is the number of **u**-free terms for z_k. If one z_k becomes more negative with the other held constant, $e^{n_k z_k/2}$ goes to zero while the other two terms in (17.41) go to finite positive constants.

The counts of **u**-free terms also help us understand why, in Figures 17.2a,b, the two legs differ in mass and width. For Classifications A and B, z_1 has more **u**-free terms than z_2 so in both contour plots the "leg" parallel to the z_2 axis is both higher and narrower than the "leg" parallel to the z_1 axis. For Classification C, z_1 has no free terms so there is no "leg" parallel to the z_2 axis.

17.2.2 Expedient 2: Ignore the Non-zero Off-Diagonals

I developed this expedient as yet another attempt to get some leverage on Michael Lavine's dynamic linear model puzzle (Chapters 6, 9, and 12). I've only used it on these DLMs but it takes advantage of features that are present in other time series and spatial models so it may be useful for other problems.

I present this expedient using Model 2, which includes signal, mystery (one harmonic), respiration (one harmonic), and heartbeat components having smoothing variances σ^2_{ss}, σ^2_{sm}, σ^2_{sr}, and σ^2_{sh} respectively, with error variance σ^2_e. (In earlier chapters I used the symbol W instead of σ^2.) Model 2's fit was a puzzle because it shrank the signal component to a straight line and shrank the respiration component close to a sine curve while the mystery curve captured almost all the variation those two components had captured in Model 1's fit.

As in Section 12.2, gather the four components' fixed effects into a single design matrix $\mathbf{X}$ and parameter β but leave their random effects in separate design matrices $\mathbf{Z}_k, k = s, m, r, h$ and random effects $\mathbf{u}_k$. Thus Model 2 for these data is

$$\mathbf{y} = \mathbf{X}\beta + \sum_k \mathbf{Z}_k \mathbf{u}_k + \varepsilon, \tag{17.42}$$

where $\mathbf{X}$ is 650×8, each $\mathbf{Z}_k$ is 650×1298, $\text{cov}(\mathbf{u}_k) = \sigma_{sk}^2 \mathbf{I}_{1298}$, and $\mathbf{R} = \sigma_e^2 \mathbf{I}_{650}$. The projection onto the orthogonal complement of $\mathbf{X}$'s column space is $\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \mathbf{K}\mathbf{K}'$ for any 650×642 matrix $\mathbf{K}$ with orthonormal columns spanning the orthogonal complement of $\mathbf{X}$'s column space. The restricted likelihood is then the likelihood arising from the transformed data $\mathbf{K}'\mathbf{y}$, which is

$$|\sum_k \sigma_{sk}^2 \mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K} + \sigma_e^2 \mathbf{I}_{642}|^{-0.5} \exp\left(-0.5\mathbf{y}'\mathbf{K}\left[\sum_k \sigma_{sk}^2 \mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K} + \sigma_e^2 \mathbf{I}_{642}\right]^{-1} \mathbf{K}'\mathbf{y}\right). \tag{17.43}$$

To put this restricted likelihood in the desired simple form, the four $\mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K}$ and $\mathbf{I}_{642}$ must be simultaneously diagonalized and that is impossible. In earlier chapters, however, we noted that the canonical re-expressions of quasi-cyclic DLM components have design matrices $\mathbf{Z}_k$ with columns that are, loosely speaking, sinusoidal with frequencies that start just above the nominal frequency of the quasi-cyclic component (e.g., 1/117 cycles per time step for the mystery term) and increase from there. Also, the signal component's canonical re-expression has design matrix columns that (loosely speaking) start with a half-cycle of a sinusoidal curve and increase in frequency from there. This suggests that if we transform the data to diagonalize $\mathbf{I}_{642}$ and $\mathbf{K}'\mathbf{Z}_s\mathbf{Z}_s'\mathbf{K}$ (for the signal component), the $\mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K}$ for mystery, respiration, and heartbeat will have mostly small off-diagonals. If so, we can approximate the restricted likelihood (17.43) by ignoring those off-diagonals and this approximation has the desired form. Because the four model components have different nominal frequencies, the components should have large diagonal elements for different canonical predictors. As we'll now see, this worked for the present problem, more or less.

So let $\mathbf{K}'\mathbf{Z}_s\mathbf{Z}_s'\mathbf{K} = \Gamma\mathbf{A}_s\Gamma'$ be the spectral decomposition of $\mathbf{K}'\mathbf{Z}_s\mathbf{Z}_s'\mathbf{K}$. Then (17.43) is proportional to

$$|\sum_k \sigma_{sk}^2 \mathbf{A}_k + \sigma_e^2 \mathbf{I}_{642}|^{-0.5} \exp\left(-0.5\mathbf{y}'\mathbf{K}\Gamma\left[\sum_k \sigma_{sk}^2 \mathbf{A}_k + \sigma_e^2 \mathbf{I}_{642}\right]^{-1} \Gamma'\mathbf{K}'\mathbf{y}\right), \tag{17.44}$$

where $\mathbf{A}_k = \Gamma'\mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K}\Gamma, k = s, m, r, h$. $\mathbf{A}_s$ is diagonal; $\mathbf{A}_m$, $\mathbf{A}_r$, and $\mathbf{A}_h$ are not, but we now approximate each one using its diagonal. Define a_{jk} to be the j^{th} diagonal element of $\mathbf{A}_k$; the approximate restricted likelihood is then

$$\prod_j (\sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2)^{-0.5} \exp\left(-0.5 \sum_j \hat{v}_j^2 / (\sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2)\right), \tag{17.45}$$

where the canonical observations $\hat{v}_j$ are $\hat{\mathbf{v}} = \Gamma'\mathbf{K}'\mathbf{y}$. This approximate restricted likelihood is identical to the likelihood from a gamma-errors GLM with identity link, "observations" $\hat{v}_j^2$ with expected value $\sum_k \sigma_{sk}^2 a_{jk} + \sigma_e^2$, and shape parameter $\nu_j = 0.5$.

This approximate restricted likelihood is also the exact restricted likelihood for an approximation to the DLM in equation (17.42), namely

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{K}\Gamma\mathbf{v} + \varepsilon \tag{17.46}$$

where $\mathbf{v}$ is $(n-8)\times 1$ with diagonal covariance $\mathbf{G}$ having $\text{var}(v_j) = \sum_{k=s,m,r,h} a_{jk}\sigma_{sk}^2$. For the signal component, $\mathbf{K}\Gamma\text{diag}(a_{js})\Gamma'\mathbf{K}' = \mathbf{K}\mathbf{K}'\mathbf{Z}_s\mathbf{Z}_s'\mathbf{K}\mathbf{K}'$ by construction. For the other components, $\mathbf{K}\Gamma\text{diag}(a_{jk})\Gamma'\mathbf{K}' \approx \mathbf{K}\mathbf{K}'\mathbf{Z}_k\mathbf{Z}_k'\mathbf{K}\mathbf{K}'$ ... we hope.

The first question about this approximation is whether the off-diagonals of the $\mathbf{A}_k$, which we're setting to zero, are in fact small. For the mystery component, almost all of them are. Considering the absolute values of the 205,761 distinct off-diagonals, the mystery term's off diagonals have 99.9$^{\text{th}}$ percentile 0.064; the 99.99$^{\text{th}}$ is 0.46 and the maximum is 0.72. For the respiration component, the off-diagonals are not quite so small. The 95$^{\text{th}}$ percentile is 0.060, the 99$^{\text{th}}$ is 0.32, and the maximum is 0.80. For the heartbeat component, the median absolute off-diagonal is 5×10^{-6}, but the percentiles increase quickly from there. The 55$^{\text{th}}$ percentile is 0.11, the 75$^{\text{th}}$ is 0.46, the 95$^{\text{th}}$ is 0.71, and the maximum is 0.82. None of this is surprising given that in the canonical representation of individual quasi-cyclic components, the progression of design-matrix columns through increasingly high frequencies is good for lower frequencies but breaks down after a point.

With this modest rationale, we'll use this approximation for what it's worth.

For two-variance models, the a_j and canonical predictors were the key to understanding the restricted likelihood. For Model 2's approximate restricted likelihood, each smoothing variance has its own a_{jk}, while there is a single set of canonical predictors, the columns of $\mathbf{K}\Gamma$ as in equation (17.46). Figure 17.3 shows the $\log(a_{jk})$ for Model 2's four components. Each component's a_{jk} vary over many orders of magnitude. The largest a_{jk} are about the same for the three quasi-cyclic components (mystery, respiration, heartbeat), while signal's largest a_{jk} are higher by about 10 logs. Signal's smallest a_{jk} are also smaller than the other components'. (This is consistent with our earlier observation that the signal component is a penalized spline and each quasi-cyclic component is similar to the ICAR model.) The peaks of the four curves follow the expected order: Signal's a_{js} are maximized at $j = 1$ by construction, mystery's a_{jm} peak at a slightly higher frequency (i.e., slightly larger j), respiration's a_{jr} peak at a somewhat higher frequency (about $j = 70$), and heartbeat's a_{jh} peak far from the others. Each curve is fairly smooth, with progressively more jiggle for mystery, respiration, and heartbeat.

As for the canonical predictors, the columns of $\mathbf{K}\Gamma$ in (17.46), recall that these are roughly sinusoidal with frequencies that increase with j. I'll describe the canonical predictors for the four groups of j for which one of the components has large a_{jk}. Figure 17.4 shows the canonical predictors for $j = 1,2,3,4$, for which the signal component's a_{js} are largest. Broadly, they are sinusoidal with 1, 1.5, 2, and 2.5 cycles, i.e., roughly quadratic, cubic, quartic, and quintic, as might be expected given

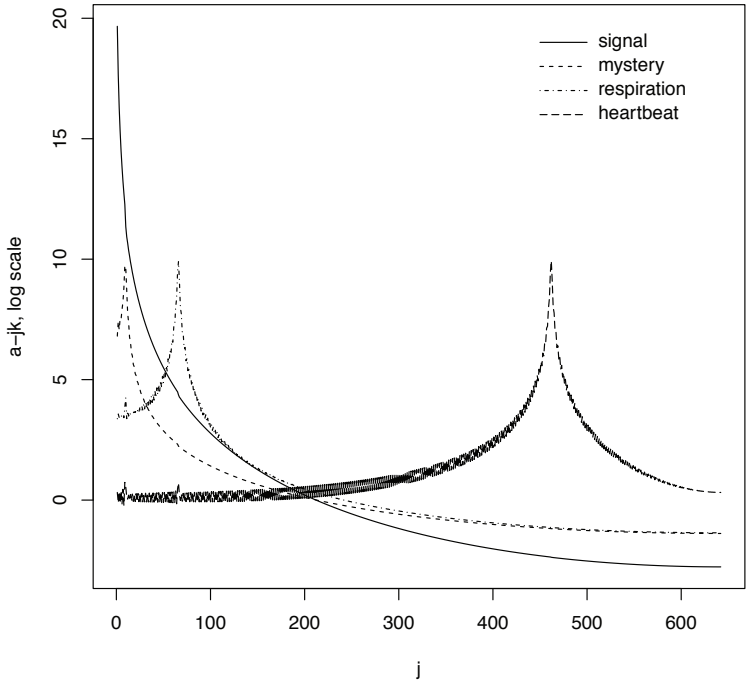

Figure 17.3: Optical imaging data, Model 2: $\log(a_{jk})$ for components k = signal, mystery, respiration, heartbeat.

that signal's fixed effects are the intercept and a linear term. These canonical predictors deviate from sinusoidal or polynomial shapes because they have been made orthogonal to the DLM's fixed effects, which include sine/cosine pairs at the nominal frequency of each quasi-cyclic component. Thus, for $j = 1$, there are six secondary peaks at the mystery component's nominal period of 117 time steps and tertiary peaks at the respiration component's nominal period.

For the three quasi-cyclic components, I describe the canonical predictors for j having large a_{jk} using their dominant period or wavelength, i.e., the reciprocal of the frequency with the greatest power in the canonical predictor's spectral density (estimated by the R function "spectrum"). Each component's largest a_{jk} occur for canonical predictors j having dominant period close to the component's nominal period. The mystery component, with nominal period 117, has largest a_{jm} for $j = 7,\ldots,13$, which have dominant periods 168.8, 135, 135, 96.4, 96.4, 96.4, and 84.4. Note that this sequence has a gap at period 117; this period was removed from the canonical predictors when they were made orthogonal to the fixed effects. Similarly, the respiration component, with nominal period 18.75, has largest a_{jr} for $j = 64,\ldots,68$, which have dominant periods 19.3, 19.3, 18.2, 18.2, and 17.8; again, note the gap at period 18.75. Finally, the heartbeat component, with nominal period 2.78, has largest

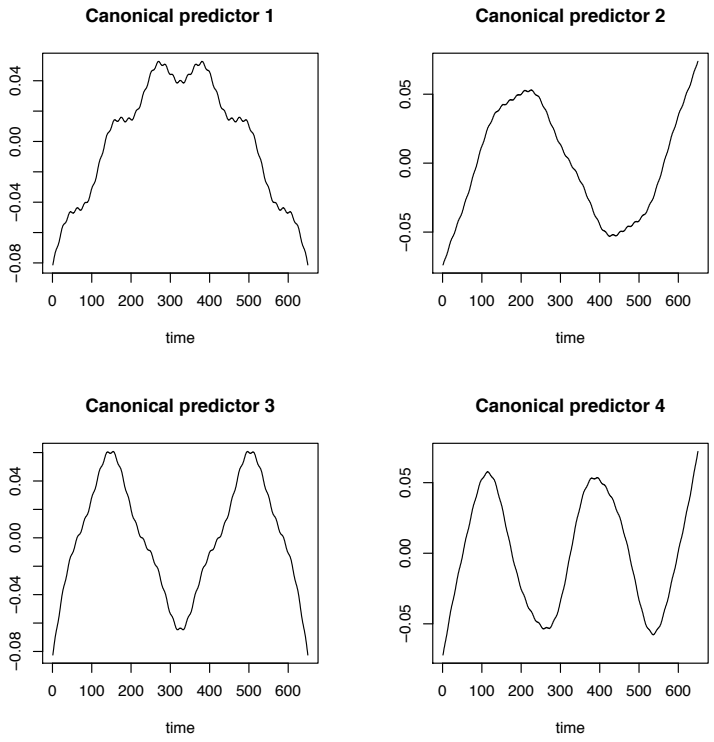

Figure 17.4: Optical imaging data, Model 2: Canonical predictors for $j = 1,2,3,4$, for which the signal component's a_{js} are largest.

a_{jh} for $j = 459,\ldots,463$, which have dominant periods 2.80, 2.80, 2.789, 2.77, 2.77, again with a gap at 2.78.

The rest of this section is as follows. First we consider the fit of Model 1, which did not have the mystery component. Then we consider Model 2, which added the mystery component, and identify features of the data and model that produced the DLM mystery. This suggests some changes to Model 2; after these changes, the modified Model 2 gives a reasonable smooth for the signal component.

So consider Model 1. The mixed-linear-model version of Model 1 is equation (17.42) but omitting the mystery component's two columns from $\mathbf{X}$ and its random effect $\mathbf{Z}_m\mathbf{u}_m$. Following the same sequence of steps as above yields an approximate restricted likelihood for Model 1 with the desired simple form. A plot of Model 1's a_{jk} looks nearly identical to Figure 17.3 but without the curve describing mystery's a_{jm}. Model 1's canonical predictors are similar to Model 2's, differing in two ways. First, Model 1 has two more canonical predictors with dominant periods about 117. Second, Model 1's canonical predictors have not been made orthogonal to a sine/cosine pair with period 117. Thus, for example, the canonical predictor for $j = 1$ lacks the six secondary peaks visible in Figure 17.4.

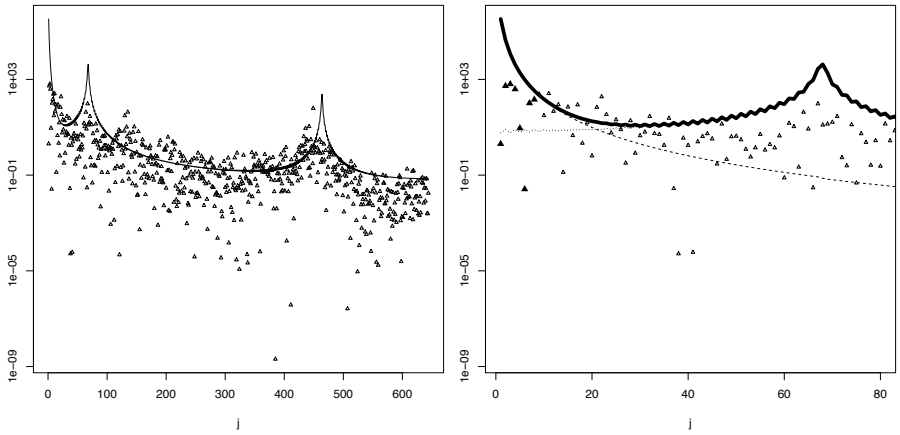

Figure 17.5: Optical-imaging data, Model 1: $\hat{v}_j^2$ (triangles) and fitted values (line) from maximizing the approximate restricted likelihood. The vertical scale is logarithmic. The left panel shows all j, the right panel shows $j = 1,\ldots,80$, with dashed and dotted lines for the signal and respiration parts of the fit, respectively.

Figure 17.5 uses these tools to examine Model 1's fit. In each panel of Figure 17.5, the horizontal axis is j, the vertical axis is logarithmic, and the triangles are the $\hat{v}_j^2$. The solid line is the fitted values $\sum_{s,r,h} \hat{\sigma}_{sk}^2 a_{jk} + \hat{\sigma}_e^2$, which have three peaks: from left to right, these arise from the signal, respiration, and heartbeat components. In the right panel, the dashed and dotted lines are the signal and respiration components' contributions to the fit respectively. If Model 1 fit these $\hat{v}_j^2$, the solid line would pass through the cloud of triangles, not through its middle, because the vertical scale is logarithmic, but a bit below its surface, as it does for j about 200 to 400.

Model 1 fits these $\hat{v}_j^2$ poorly. The signal component's fit creates a peak that is larger than $\hat{v}_j^2, j = 1,\ldots,8$, which are shown as solid triangles in Figure 17.5's right panel. The respiration component's fit creates a peak for j around 50-80, which is also too high. Further, the $\hat{v}_j^2$ show peaks at about 140 and arguably 210, which the fit cannot capature. Thus Model 1's fit is a compromise: the variances for signal and respiration, $\hat{\sigma}_{ss}^2$ and $\hat{\sigma}_{sr}^2$, are pulled up by $\hat{v}_j^2$ for j about 10-30 and $\hat{\sigma}_{sr}^2$ is also pulled by the peaks in the $\hat{v}_j^2$ at about 140 and 210. These latter two peaks correspond to frequencies about 2 and 3 times respiration's nominal frequency, i.e., its second and third harmonics; below I call them respiration's "echo" peaks.

The mystery component was added to Model 1 to capture variation with frequencies between those of the signal and respiration components, i.e., to help fit $\hat{v}_j^2$ for j about 5-30. Figure 17.6 shows the results. (Recall that the $\hat{v}_j^2$, a_{jk}, and canonical predictors are not identical for Models 1 and 2, though they are quite similar.) The fitted values again have three peaks but now arising, from left to right, from mystery,

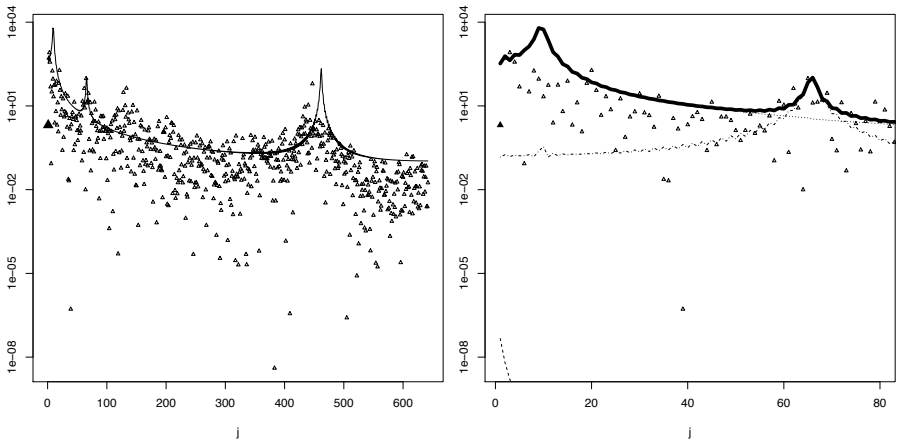

Figure 17.6: Optical-imaging data, Model 2: $\hat{v}_j^2$ (triangles) and fitted values (line) from maximizing the approximate restricted likelihood. The vertical scale is logarithmic. The left panel shows all j, the right panel shows $j = 1, \ldots, 80$, with dashed, dotted, and dash-dot lines for the signal, mystery, and respiration parts of the fit, respectively.

respiration, and heartbeat. Figure 17.6's right panel shows that the contribution of signal to this fit is negligible (it is barely visible in the panel's lower left corner). The peak arising from respiration, for j about 70, now fits better because the mystery component fits smaller j. Overall, however, this fit is still poor. The mystery component creates a peak in the fitted values where the $\hat{v}_j^2$ have no corresponding peak, and the two "echo" peaks for j around 140 and 210 are still ignored.

The signal component is effectively reduced to its linear fixed effects because with the mystery component included in the model, the first few j are very influential for $\hat{\sigma}_{ss}^2$, and $\hat{v}_1^2$ (the solid triangle in Figure 17.6) is quite small. The influence of this and other $\hat{v}_j^2$ can be seen in Figure 17.7, which shows scaled residuals and one-step approximate changes in the variance estimates from omitting individual $\hat{v}_j^2$, computed as in Chapter 16. Although $\hat{v}_1^2$ has a small scaled residual, it is extremely influential: removing $\hat{v}_1^2$ makes $\hat{\sigma}_{sm}^2$ much smaller and $\hat{\sigma}_{ss}^2$ and $\hat{\sigma}_{sr}^2$ much larger.

The two "echo" peaks, $\hat{v}_j^2$ for j around 140 and 210, are also influential: removing $\hat{v}_j^2$ in either of these peaks, but especially around $j = 140$, reduces $\hat{\sigma}_{sm}^2$ and increases $\hat{\sigma}_{sr}^2$. This is odd but true: In Figure 17.6's right panel, mystery's contribution to the fit (shown as a dotted line) actually exceeds respiration's contribution (shown as a dash-dot line) for j larger than about 70. That is, mystery's variance σ_{sm}^2 is inflated to capture part of the "echo" peaks, even though respiration's a_{jr} are large for j closer to the "echo" peaks than are mystery's a_{jm}.

(The heartbeat component is off in its own world, driven by j around 400-500 and unaffected by the foregoing.)

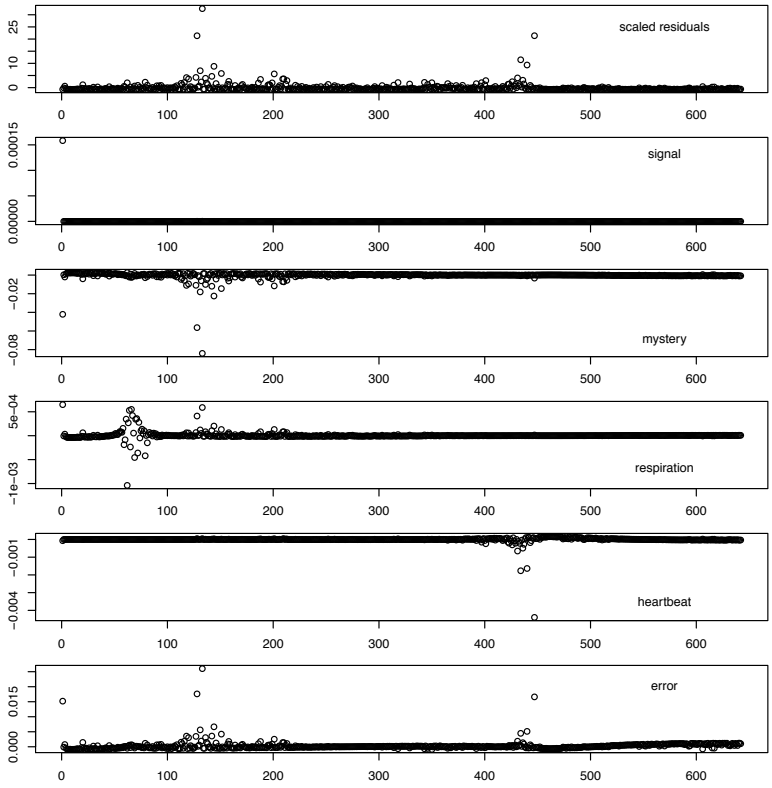

Figure 17.7: Optical imaging data, Model 2: Scaled residuals (top panel) and approximate changes in each variance estimate (next five panels) from deleting each individual $\hat{v}_j^2$. In all plots, the horizontal axis is j.

It appears that lack of fit to signal and respiration forces mystery's σ_{sm}^2 upward to "fix" the bad fit, which in turn pushes down σ_{ss}^2 and σ_{sr}^2 and produces the idiotic smoothed signal. The foregoing suggests two possible solutions: Add second and third harmonics for respiration to capture the two "echo" peaks, or omit $\hat{v}_1^2$. Regarding the latter, in Chapter 16 we could do that simply by moving a column from the random-effect design matrix to the fixed-effect design matrix, but we cannot do the same for this DLM because $\hat{v}_1^2$ is defined only for the approximate restricted likelihood and has no corresponding column in any exact model's design matrix. We *can* however do something that has almost the same effect: replace the signal component's locally linear model with a locally quadratic model. With this model, the signal component has three columns in the fixed-effect design matrix $\mathbf{X}$, for an intercept and linear and quadratic terms, and the canonical predictor for $j = 1$ is now roughly cubic instead of quadratic. Thus Model 2's troublesome $\hat{v}_1^2$ goes away.

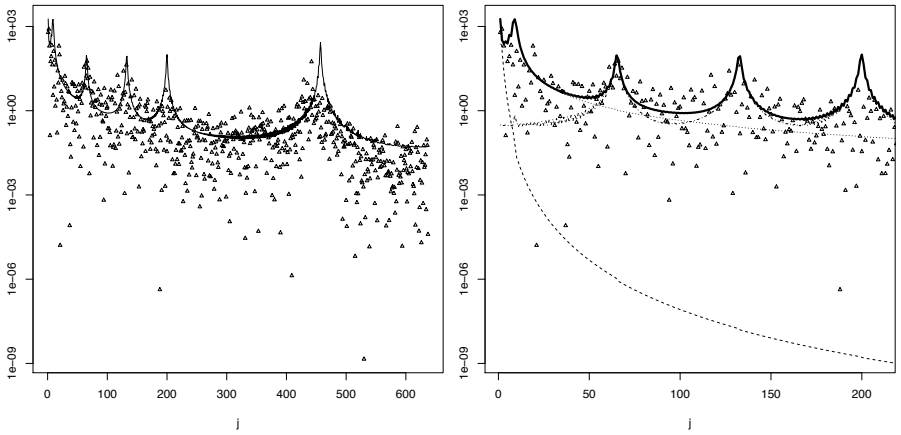

Figure 17.8: Optical-imaging data, Model 2 changed to include a locally-quadratic model for signal and three harmonics for respiration: $\hat{v}_j^2$ (triangles) and fitted values (line) from maximizing the approximate restricted likelihood. The vertical scale is logarithmic; the left panel shows all j, while the right panel shows $j = 1, \ldots, 210$. Right panel: dashed, dotted, and dash-dot lines are the signal, mystery, and respiration parts of the fit, respectively.

It turns out that changing Model 2 in either of these ways individually, or both ways together, produces the desired result. For Model 1 (with either or both of these changes), the fitted signal still has the low-frequency junk that motivated Model 2, but Model 2 (as modified) now gives a sensible signal fit. Figure 17.8 shows the $\hat{v}_j^2$ and the fit for Model 2 with both changes. A plot analogous to Figure 17.7 (which I haven't included) shows that the fitted σ_{ss}^2 is sensitive to $\hat{v}_1^2$ and $\hat{v}_2^2$ but now mystery and respiration are much less so. In Figure 17.8, this new fit still has a peak arising from the mystery component that doesn't fit anything in the $\hat{v}_j^2$, but at least it's smaller than for earlier models. Also, while this model fits the peaks around $j = 140$ and 210 better than earlier models, it is too high around $j = 210$.

Alas, now I must complicate this happy ending by pointing out two things I've been sweeping under the rug.[2] First, as a student asked me, does this make sense as a way to do data analysis? You could ask the same question about regression diagnostics, but it *is* a reasonable question. This section's purpose was to understand why these DLMs fit the way they do, in particular how lack of fit affects variance estimates. It pursued that purpose by devising diagnostics to explore lack of fit in the variance part of the model and suggest models that fit better. Although I arguably manipulated the analysis to a pre-determined conclusion, I am comfortable with that because a plot of this data series plainly shows the general shape the smoothed signal must have. I agree however that mimicking what I've done would be problematic as a

[2]Much of the following is based on work done with Ellie Duffy after the book was published.

general principle, especially if the object was to extract a small signal. At this point, though, we are so desperate for tools that the gain in general understanding seems worth that bit of discomfort.

The other problem I've been concealing is that these DLMs are extremely fussy to fit. For this and several other series from Michael Lavine's dataset, my research assistant Ellie Duffy and I used the dlm package in R (Petris 2010) to maximize the exact likelihood, then tried to maximize the exact restricted likelihood using our own (naïve) code, then tried to maximize the approximate restricted likelihood. In all these analyses, we parameterized the model using log variances, which seems necessary because the variances range over several orders of magnitude. However, this makes it impossible to get a zero variance estimate and it seems clear that the error variance should be estimated at zero and, for some models, so should the signal variance. The exact likelihood and restricted likelihood and the approximate restricted likelihood are very flat with respect to logs of variances that probably should have zero estimates. For some of these data series, the approximate restricted likelihood appears to have many local maxima; so far, we have not found more than two local maxima for the likelihood or exact restricted likelihood. For some models, however, we have been unable to maximize the exact restricted likelihood.

The original version of this section compared the exact and approximate restricted likelihoods for Model 2 applied to Michael Lavine's series 1 (the only series for which I have presented results). The agreement between these two functions was, in light of later experience, remarkably good: both functions had two local maxima, occurring for very similar values of the five variances, and having very similar function values. The fitted values shown above used the (apparently) global maximum of the approximate restricted likelihood but would have been nearly identical using the maximum of the exact restricted likelihood. This is not always the case with other series; I was outlandishly lucky with my first example.

Perhaps the moral of the story is that this model is too complex for a data series of length 650. This points to a related puzzle: why is the error variance always zero? In all the models I've fit to this series, with many starting values, I have only rarely found even a secondary maximum at which σ_e^2 was not effectively zero. It appears this DLM differs from Chapter 16's two-variance models in that for all j, some non-error component competes with error to explain $\hat{v}_j^2$, i.e., no range of j is distinctively informative about σ_e^2. Based on Figure 17.3, I hypothesize that the heartbeat component competes with error because heartbeat's $a_{jh} > 1$ for large j, unlike the other components' a_{jk}. In a variant of Model 2 omitting the heartbeat component, error absorbed variation previously captured by heartbeat, while the rest of the model's fit was essentially unchanged. If the data series were much longer, the a_{jk} for all components would become small enough for the largest j so that those $\hat{v}_j^2$ would provide information about σ_e^2. Nothing, however, can alleviate the confounding of signal and mystery: signal's fit is necessarily sensitive to the first few $\hat{v}_j^2$.