Oddity #1: Adding a spatial RE wipes out a clear association

A lot is known about the ***mechanics*** of what's happening ("spatial confounding").

We know how to alter the spatial random effect so it doesn't happen.

There's now a debate underway about

- how to interpret spatial confounding, and
- what, if anything, to do about it.

I'll summarize what's in Hodges (2013) and some recent developments.

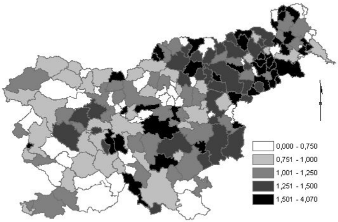
Recap of Oddity #1

Dr. Vesna Zadnik was interested in the association of stomach cancer with socioeconomic status in Slovenia.

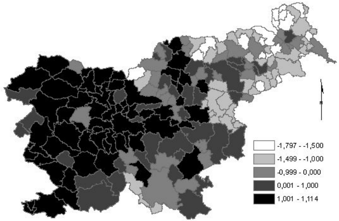
Dataset: For the $i = 1, \ldots, 194$ municipalities that partition Slovenia

- O_i is the observed count of stomach cancer cases
- E_i is the expected count using indirect standardization
- SEc_i is the centered socioeconomic status (SES) score

Outcome: $SIR_i = O_i / E_i$



Predictor SEc_i.



First, a non-spatial model

Dr. Zadnik first did a non-spatial analysis:

$O_i \sim$ Poisson with $\log\{E(O_i)\} = \log(E_i) + \alpha + \beta SEc_i$,

with flat priors on α and β.

This analysis gave the obvious result: $\beta|\{O_i\}$ had

- median -0.14
- 95% interval $(-0.17, -0.10)$.

This result captures the negative association that's obvious in the plots.

Now, a spatial analysis

Object: Discount the sample size to account for spatial correlation.
(Other people have different objectives, as we'll see.)

$O_i \sim$ Poisson with $\log\{E(O_i)\} = \log(E_i) + \beta SEc_i + S_i + H_i$,

This has two intercepts:

- Spatial similarity: $S_i \sim L_2$-norm ICAR, precision τ_s.
- Heterogeneity: $H_i \sim$ iid Normal, mean zero, precision τ_h.

Priors:

- independent gammas for τ_h and τ_s, mean 1 and variance 100,
- flat prior for β.

SURPRISE!

	DIC	p_D	β's median	β's 95% interval
Non-spatial model	1153	2	-0.14	(-0.17, -0.10)
Spatial model	1082	62	-0.02	(-0.10, 0.06)

After adding the spatial and heterogeneity random effects:

- β's posterior SD increased, which we expected
- β's posterior median became effectively zero, which we didn't.

Adding the spatial random effect makes an obvious association go away.

Why?

Apparently faulty analogy: In GEE analyses, in my [previous] experience, you needed a huge within-cluster correlation to affect point estimates.

Mechanics of spatial confounding – Hodges & Reich (2010)

Let's consider the normal-errors version of this model:

For n-dimensional $\mathbf{y}$, model $\mathbf{y} = \mathbf{F}\beta + \mathbf{I}_n\mathbf{S} + \epsilon$:

- $\mathbf{y}$, $\mathbf{F}$, $\mathbf{S}$, and ϵ are $n \times 1$, β is scalar;
- $\mathbf{F}$ is centered and scaled;
- $\mathbf{S} \sim$ ICAR with precision parameter τ_s, neighbor matrix $\mathbf{Q}$;
- $\epsilon \sim N_n(\mathbf{0}, \frac{1}{\tau_e}\mathbf{I})$, where $\tau_e = 1/\sigma_e^2$

The intercept is implicit in the ICAR model for $\mathbf{S}$.

Obvious concern: A regression with n observations and n+p predictors.

The potential for collinearity is obvious.

To clarify this, let's re-parameterize the random effect $\mathbf{S}$.

Spatial confounding: re-parameterizing $\mathbf{S}$

Original: $\mathbf{y} = \mathbf{F}\beta + \mathbf{I}_n\mathbf{S} + \epsilon$, $\mathbf{S} \sim \text{ICAR}(\tau_s, \mathbf{Q})$.

Re-parameterized: $\mathbf{y} = \mathbf{F}\beta + \mathbf{V}\mathbf{b} + \epsilon$, where

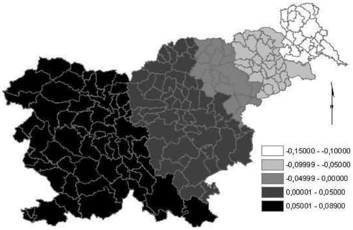
- $\mathbf{Q} = \mathbf{V}\mathbf{D}\mathbf{V}'$ is $\mathbf{Q}$'s spectral decomposition
- $\mathbf{V}$ is orthogonal, so $\mathbf{V}\mathbf{V}' = \mathbf{V}'\mathbf{V} = \mathbf{I}_n$
- $\mathbf{D}$ is diagonal, $d_1 \geq d_2 \geq \ldots d_{n-I} > 0 = d_{n-I+1} = \cdots = d_n$
- $\mathbf{b} \sim$ Normal, mean $\mathbf{0}$, precision $\tau_s\mathbf{D}$; $\mathbf{b} = \mathbf{V}'\mathbf{S}$.

Why does this help? τ_s controls shrinkage of $b_i, i \leq n - I$, and:

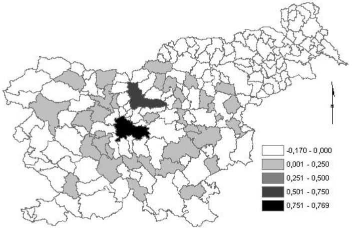
- $\mathbf{b}$'s elements are smoothed independently of each other
- b_1 is smoothed most (d_1 is largest)
- b_{n-I} is smoothed least (d_1 is the smallest positive d_j)
- $b_{n-I+1}, \ldots, b_n$ aren't smoothed at all.

and $\mathbf{V}$'s columns are interpretable.

Eigenvector for $d_{193} = 0.03$, roughly east-west gradient.



Eigenvector for $d_1 = 14.46$, regions with most neighbors vs. ...



Spatial confounding in linear-model terms

Without the ICAR random effect, $\mathbf{y} \sim N(\mathbf{1}_n\alpha + \mathbf{F}\beta, \mathbf{I}/\tau_e)$, and

$$E(\beta|\mathbf{y}, \tau_e) = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y} \qquad \text{flat prior on } \beta$$

With the ICAR random efect, $\mathbf{y} \sim N(\mathbf{F}\beta + \mathbf{V}\mathbf{b}, \mathbf{I}/\tau_e), \mathbf{b}$ has precision $\tau_s\mathbf{D}$

$$E(\beta|\tau_e, \tau_s, \mathbf{y}) = (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{y} - (\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y}),$$

where $E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y})$ is *not* conditional on β

$$E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y}) = (\mathbf{V}'\mathbf{P}^c\mathbf{V} + r\mathbf{D})^{-1}\mathbf{V}'\mathbf{P}^c\mathbf{y},$$

for $r = \tau_s/\tau_e = \sigma_e^2/\sigma_s^2$ and $\mathbf{P}^c = \mathbf{I} - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$.

The change from adding the spatial RE is $-(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'\mathbf{V}E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y})$.

When does spatial confounding occur?

Change from adding the spatial RE: $\quad -(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' \quad \mathbf{V}E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y})$.

$\mathbf{V}$ is orthogonal and $\mathbf{F}$ is centered and scaled ($\mathbf{F}'\mathbf{F} = n-1$), so

correlation$(\mathbf{F}, \mathbf{V}) = (n-1)^{-0.5}\mathbf{V}'\mathbf{F}$.

$\Rightarrow E(\beta|\tau_e, \tau_s, \mathbf{y}) = \hat{\beta}_{OLS} - (n-1)^{1/2}\boldsymbol{\rho}' E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y})$.

Spatial confounding will occur when:

1. $\mathbf{F}$ is highly correlated with $\mathbf{V}_j$, the j^{th} column of $\mathbf{V}$,
2. $\mathbf{y}$ is correlated with $\mathbf{V}_j$ after accounting for $\mathbf{F}$,
3. d_j is small.

(1) and (2) define confounding in linear models;

(3) implies b_j is not shrunk much.

Spatial confounding is just collinearity

Spatial confounding occurs in the Slovenia data because, for $j = 193$:

1. $\mathbf{F}$ is highly correlated with $\mathbf{V}_{193}$, the 193^{rd} column of $\mathbf{V}$.
    - Correlation$(SEc, V_{193}) = 0.72$
2. $\mathbf{y}$ is correlated with both $\mathbf{F}$ and $\mathbf{V}_{193}$.
    - This is visible in the plot of SIR.
3. d_j is small.
    - $d_{193} = 0.03$, the smallest positive eigenvalue.

$\mathbf{S}$ is not smoothed much: the effective number of parameters is $p_D = 62.3$.

In linear model terms, the effect on β's estimate is the effect of adding a collinear regressor to a linear model. (Variance inflation occurs, too.)

The mechanics, in a more spatial-statistics style

Rewrite the model as $\mathbf{y} = \mathbf{F}\beta + \psi$, where $\psi = \mathbf{Vb} + \epsilon$; assume $I = 1$.

$\mathbf{S} = \mathbf{Vb}$ has a singular covariance matrix, so we proceed indirectly.

Partition $\mathbf{V} = (\mathbf{V}^{(1)}|\mathbf{V}^{(2)})$, so $\mathbf{V}^{(1)}$ is $n \times (n - I)$, $\mathbf{V}^{(2)}$ is $n \times I$.
Partition $\mathbf{b} = (\mathbf{b}^{(1)'}|\mathbf{b}^{(2)\prime})'$, so $\mathbf{b}^{(1)}$ is $(n - I) \times 1$, $\mathbf{b}^{(2)}$ is $I \times 1$.

Pre-multiply the model equation by $\mathbf{V}'$, so it becomes

$$\begin{aligned} \mathbf{V}^{(1)'}\mathbf{y} &= \mathbf{V}^{(1)'}\mathbf{F}\beta + \mathbf{e}_1, & \text{prec}(\mathbf{e}_{1i}) = \tau_e(rd_i)/(1 + rd_i) < \tau_e \\ \mathbf{V}^{(2)'}\mathbf{y} &= \mathbf{b}^{(2)} + \mathbf{e}_2, & \text{prec}(\mathbf{e}_{2i}) = \tau_e \end{aligned}$$

where $\mathbf{e}_1 = \mathbf{b}^{(1)} + \mathbf{V}^{(1)'}\epsilon$ and $\mathbf{e}_2 = \mathbf{V}^{(2)'}\epsilon$, $r = \tau_s/\tau_e = \sigma_e^2/\sigma_s^2$.

$\mathbf{V}^{(2)} \propto \mathbf{1}_n$ and $\mathbf{V}^{(2)'}\mathbf{F} = 0$ because $\mathbf{F}$ is centered.

All the information about β comes from $\mathbf{V}^{(1)'}\mathbf{y}$.

The transformed model equation becomes

$$\begin{aligned}
\mathbf{V}^{(1)'}\mathbf{y} &= \mathbf{V}^{(1)'}\mathbf{F}\beta + \mathbf{e}_1, \qquad \text{prec}(\mathbf{e}_{1i}) = \tau_e(rd_i)/(1+rd_i) < \tau_e \\
\mathbf{V}^{(2)'}\mathbf{y} &= \mathbf{b}^{(2)} + \mathbf{e}_2, \qquad \text{prec}(\mathbf{e}_{2i}) = \tau_e
\end{aligned}$$

where $\mathbf{e}_1 = \mathbf{b}^{(1)} + \mathbf{V}^{(1)'}\epsilon$ and $\mathbf{e}_2 = \mathbf{V}^{(2)'}\epsilon$, $r = \tau_s/\tau_e = \sigma_e^2/\sigma_s^2$.

<u>Without</u> **S**: $\text{prec}(\beta|\mathbf{y},\tau_e) = \mathbf{F}'\mathbf{F}\tau_e = (n-1)\tau_e$

<u>With</u> **S**: $\text{prec}(\beta|\mathbf{y},\tau_s,\tau_e) = (n-1)\tau_e - (n-1)\tau_e \sum_{i=1}^{n-1} \rho_i^2/(1+rd_i)$,

$\rho_i = \text{correlation}(\mathbf{F}, V_i) = (n-1)^{-1/2} V_i'\mathbf{F}$.

Information loss is large if: r is small <u>and</u> ρ_i is large for i with small d_i.

Different rows of $\mathbf{V}^{(1)'}\mathbf{y}$ have different information about $\beta \Rightarrow$

if row i is effectively deleted by large ρ_i & small d_i, then
$E(\beta|\mathbf{y})$ can change a lot after adding **S** to the model.

Avoiding spatial confounding: restricted spatial regression

Idea: Attribute to **F** all the variation in **y** that **F** and **S** compete over.

Simplest (not necessarily best) way to do this:

Replace $\mathbf{y} = \mathbf{F}\beta + \mathbf{I}_n\mathbf{S} + \boldsymbol{\epsilon}$ with $\mathbf{y} = \mathbf{F}\beta + \mathbf{P}^c\mathbf{S} + \boldsymbol{\epsilon}$.

where $\mathbf{P}^c = \mathbf{I}_n - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}'$,
the residual projection for a regression on **F**.

Conditional on (τ_s, τ_e), β has the same posterior mean as in the analysis without **S**, but larger posterior variance (as it should).

How to interpret spatial confounding? What to do?

The answer depends on whether **S** is an old-style or new-style RE.

- **S is a new-style RE, a formal device to implement a smoother**
  - (i) Spatially correlated errors remove bias in estimating β and are generally conservative.
  - (ii) Spatially correlated errors can introduce or remove bias in estimating β and are not necessarily conservative.
- **S is an old-style RE**
  - (iii) The regressors **V** implicit in the spatial effect **S** are collinear with the fixed effect **F**, but neither estimate of β is biased.
  - (iv) Adding the spatial effect **S** creates information loss, but neither estimate of β is biased.
  - (v) Because error is correlated with the regressor **F** in the sense commonly used in econometrics, *both* estimates of β are biased.

Except for (v), these interpretations treat **F** as measured without error and not drawn from a probability distribution.

$\mathbf{S}$ is just a formal device to implement a smoother

"Adding spatially correlated errors adjusts $\hat{\beta}$ for spatially structured missing covariates" even if we don't know what those covariates are.

Does it?

Suppose data arise from $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \mathbf{H}\gamma + \boldsymbol{\epsilon}$, where $\mathbf{H}$ is centered and scaled.

Suppose we fit these three models to the data:

- **Model 0**: $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \boldsymbol{\epsilon}$
- **Model H**: $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \mathbf{H}\gamma + \boldsymbol{\epsilon}$ $\leftarrow$ correct model
- **Model S**: $\mathbf{y} = \mathbf{F}\beta + \mathbf{V}\mathbf{b} + \boldsymbol{\epsilon}$, $\leftarrow$ "hope it works!"

where $\mathbf{S} = \mathbf{V}\mathbf{b}$ is modeled as an ICAR.

It's not that hard to derive $E(\beta|\mathbf{y}, \tau_s, \tau_e)$.

Here's what you actually get

- **Model 0**: $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \boldsymbol{\epsilon}$
- **Model H**: $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \mathbf{H}\gamma + \boldsymbol{\epsilon}$ $\quad\leftarrow$ correct model
- **Model S**: $\mathbf{y} = \mathbf{F}\beta + \mathbf{V}\mathbf{b} + \boldsymbol{\epsilon}$, $\quad\leftarrow$ "hope it works!"

Conditional on $\mathbf{y}, \tau_s, \tau_e$,

$$
\begin{aligned}
\hat{\beta}^{(H)} &= \hat{\beta}^{(0)} \left[1 - \frac{\frac{\rho_{FH}\rho_{HY}}{\rho_{FY}} - \rho_{FH}^2}{1 - \rho_{FH}^2} \right], \text{ the right estimate,} \\
\hat{\beta}^{(S)} &= \hat{\beta}^{(0)} \left[1 - \frac{\frac{\rho'_{FY}}{\rho_{FY}} - q}{1 - q} \right], \text{ supposedly like } \hat{\beta}^{(H)}
\end{aligned}
$$

- ρ_{AB} = correlation(A, B)
- $\rho'_{FY} = \mathbf{F}'(I + r\mathbf{Q})^{-1}\mathbf{y}/((n-1)\mathbf{y}'\mathbf{y})^{0.5}$
- $q = \mathbf{F}'(I + r\mathbf{Q})^{-1}\mathbf{F}/(n-1)$

Comments on the previous slide's results

$\hat{\beta}^{(H)}$ and $\hat{\beta}^{(S)}$ have no necessary relation to each other.

If $\rho_{FH} = 0$, then $\hat{\beta}^{(H)} = \hat{\beta}^{(0)}$ but $\hat{\beta}^{(S)}$ can be larger or smaller than $\hat{\beta}^{(H)}$.

When $\rho'_{_{FY}} \approx \rho_{_{FY}}$, $\hat{\beta}^{(S)} \approx 0$.

This happens if there's a j such that

- the correlation of $\mathbf{F}$ and V_j is large;
- the correlation of $\mathbf{y}$ and V_j is large;
- and d_j and r are small,

  as in the Slovenian data.

Assuming $\mathbf{y} = \mathbf{1}_n\alpha + \mathbf{F}\beta + \mathbf{H}\gamma + \boldsymbol{\epsilon}$, taking the expectation of these conditional means over the distribution of $\mathbf{y}$ gives

$$
\begin{aligned}
E(\hat{\beta}^{(H)} - \hat{\beta}^{(0)} | \tau_e) &= \rho_{FH}\gamma \\
E(\hat{\beta}^{(S)} - \hat{\beta}^{(0)} | \tau_e, \tau_s) &= \left[\frac{\frac{\rho'_{FH}}{\rho_{FH}} - q}{1 - q} \right] \rho_{FH}\gamma \qquad \text{if } \rho_{FH} \neq 0 \\
&= \frac{\rho'_{FH}}{1 - q}\gamma \qquad \text{if } \rho_{FH} = 0,
\end{aligned}
$$

where $\rho'_{FH} = \mathbf{F}'(I + r\mathbf{Q})^{-1}\mathbf{H}/(n-1)$.

If $\gamma \neq 0$, the adjustment under Model S can be biased + or -.

If $\gamma = 0$, i.e., $\mathbf{H}$ doesn't matter, Model S gives an unbiased adjustment.

Interpretation (i) is wrong; interpretation (ii) is right

- (i) Spatially correlated errors remove bias in estimating β and are generally conservative.
- (ii) Spatially correlated errors can introduce or remove bias in estimating β and are not necessarily conservative.

Conclusion: Adding spatially correlated errors is not conservative:

> A canonical regressor V_i that is collinear with $\mathbf{F}$ can cause β's estimate to increase in magnitude.

In cases in which β's estimate should not be adjusted, introducing spatially correlated errors can adjust the estimate haphazardly.

S is an old-style random effect

- (iii) The regressors **V** implicit in the spatial effect **S** are collinear with the fixed effect **F**, but neither estimate of β is biased.
- (iv) Adding the spatial effect **S** creates information loss, but neither estimate of β is biased.

Assume **F** is fixed and known: Both (iii) and (iv) are correct.

Interpretation (iv) follows because GLS gives unbiased estimates even when the error covariance is specified incorrectly.

Interpretation (iii):

$$E(\beta|\tau_e, \tau_s, \mathbf{y}) = \hat{\hat{\beta}}_{OLS} - (n-1)^{-1/2}\boldsymbol{\rho}' E(\mathbf{b}|\tau_e, \tau_s, \mathbf{y}) = \hat{\hat{\beta}}_{OLS} - \mathbf{K}\mathbf{P}^c\mathbf{y},$$

and $\mathbf{P}^c = \mathbf{I} - \mathbf{F}(\mathbf{F}'\mathbf{F})^{-1}\mathbf{F}' \Rightarrow \mathbf{P}^c\mathbf{y} = \mathbf{P}^c(\mathbf{V}\mathbf{b} + \boldsymbol{\epsilon})$.

$\Rightarrow$ the spatial and OLS estimates of β have the same expectation wrt **y**, and the OLS is unbiased, so both are.

$\mathbf{S}$ is an old-style random effect

- (v) Because error is correlated with the regressor $\mathbf{F}$ in the sense commonly used in econometrics, *both* estimates of β are biased.

Now assume $\mathbf{F}$ is a random variable.

The model is $\mathbf{y} = \mathbf{F}\beta + \boldsymbol{\psi}$;
the error $\boldsymbol{\psi} = \mathbf{V}\mathbf{b} + \boldsymbol{\epsilon}$ has a non-diagonal covariance matrix.

Because $\mathbf{F}'\mathbf{V} \neq 0$, $\mathbf{F}$ is correlated with $\boldsymbol{\psi}$.

$\Rightarrow$ *both* the OLS and spatial estimates of β are biased
(standard result in econometrics).

My view until a couple years ago:

You should *always* use restricted spatial regression.

Hodges & Reich (2010 TAS) was deliberately provocative, intended to start a discussion.

A discussion has indeed started.

On the mechanics of spatial confounding

How the data determine the variances: H2013 Section 15.2.3

- In the RL, all variation in $\mathbf{y}$ that's in the column space of $\mathbf{X}$ is credited to the fixed effects.
- Thus, all the information in $\mathbf{y}$ about σ_s^2 is in $\mathbf{P}_X^c\mathbf{y}$.

Hughes & Haran (JRSSB 2013) give a better restricted spatial random effect that has:

- a more intuitive covariance matrix.
- much smaller dimension $\Rightarrow$ much faster computing.

Hanks et al (Environmetrics 2015) extended spatial confounding and RSR to random effects distributed as Gaussian processes.

On whether to use restricted spatial regression (RSR)

Mixed linear models with spatial REs are used for disparate purposes.

The choice to use RSR or not should depend on the purpose.

- *Interpolation/prediction*: Use RSR $\Leftrightarrow$ it improves a beauty measure.

G. Page on interpolation: "[T]he bias [in interpolation/prediction] depends crucially on spatial scale in [$\mathbf{y}$ and $\mathbf{F}$] ... spatial confounding can actually improve prediction performance in terms of MSPE" although RSR is better in some circumstances.

- *Causal inference*: The jury is out.

G. Papadogeorgou (spoke here last semester!): Adjust for unmeasured spatially-distributed covariates using a weighted average of

  - a conventional propensity score and
  - a measure of spatial distance.

Oddity #2: Adding a RE changes a group comparison

Recall:

- Each child (cluster) had 1-4 crowns.
- Comparisons of crown types (groups) changed a lot when we included a random effect for child (cluster).

Summary:

- Adding the RE for child changes the estimated effects only if
  - ≥ 1 contrast in a group's cluster sizes is large, and
  - the same contrast in a group's cluster averages $\bar{y}_i$ is also large.
- H2013 $\Leftrightarrow$ **X** and **Z** are collinear in certain way.

This is an instance of "informative cluster size":

- If cluster size is informative as described above,
- then design matrices **X** and **Z** are collinear in a certain way, and
- adding the RE to the model changes the estimated tx effect.

Oddity #3: Differential shrinkage of equal-sized effects

Recall:

- Smoothed ANOVA in cancer clinical trial data.
- Equal-sized unshrunk effects were shrunk to different extents.

Hypothesis: Collinearity among fixed and random effects causes this.

- Tested using normal-errors, normal-RE model with 1 FE, 2 REs.
- Hypothesis fails: collinearity *per se* did not produce differential shrinkage.

New hypotheses:

1. (Boring) Coding error or MCMC failure in the original analysis.
2. Differential shrinkage can happen in normal-errors problems but only if > 3 predictors.
3. It's caused by some aspect of the time-to-event analysis that is absent in the normal-errors analysis.

Oddity #4: Adding an RE wipes out two other REs

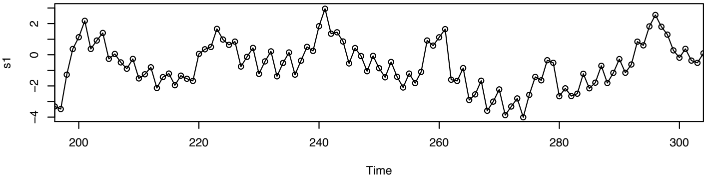
Testing a new method to localize epileptic activity (Lavine et al).

y_t = % change in average pixel value for light of wavelength 535 nm,
$t = 0, \ldots, 649$, with time steps of 0.28 sec.

Stimulus was applied during time steps $t = 75$ to 94

Object: Estimate the response to the stimulus.

Complication: artifacts from heartbeat and respiration, with periods of 2–4 and 15–25 time steps.



Model 1: Smooth response, quasi-cyclic terms for artifacts

$y_t =$ % change in average pixel value for light of wavelength 535 nm,
$t = 0, \ldots, 649$, with time steps of 0.28 sec.

Stimulus was applied during time steps $t = 75$ to 94

Model: a DLM with observation equation

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

- $s_t =$ smoothed response, the object of this analysis
- h_t, r_t are heartbeat and respiration respectively
- $v_t \sim$ iid $N(0, W_v)$.

State equations for s_t, h_t, r_t

State equation for s_t is the linear growth model:

$$\begin{pmatrix} s_t \\ \mathrm{slope}_t \end{pmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} s_{t-1} \\ \mathrm{slope}_{t-1} \end{pmatrix} + \mathbf{w}_{s,t},$$

$$\mathbf{w}'_{s,t} = (0, w_{slope,t}) \text{ and } w_{slope,t} \sim \text{ iid } N(0, W_s).$$

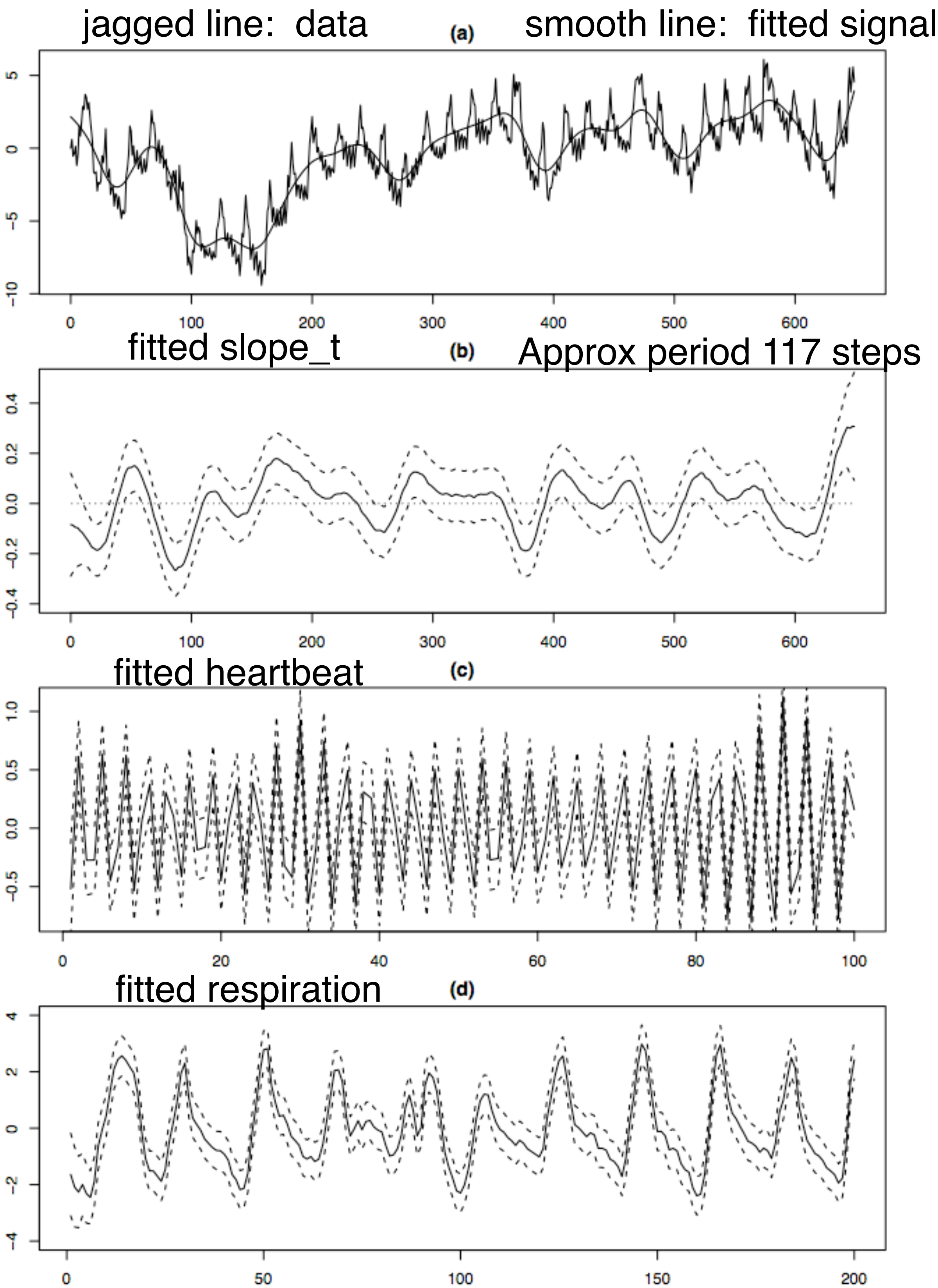
State equation for quasi-cyclic components (this is for heartbeat):

$$\begin{pmatrix} b_t \cos\alpha_t \\ b_t \sin\alpha_t \end{pmatrix} = \begin{bmatrix} \cos\delta_h & \sin\delta_h \\ -\sin\delta_h & \cos\delta_h \end{bmatrix} \begin{pmatrix} b_{t-1} \cos\alpha_{t-1} \\ b_{t-1} \sin\alpha_{t-1} \end{pmatrix} + \mathbf{w}_{h,t},$$

$$\mathbf{w}'_{h,t} = (w_{h1,t}, w_{h2,t}) \sim \text{ iid } N_2(0, \mathbf{W}_h) \text{ for } \mathbf{W}_h = W_h \mathbf{I}_2.$$

Periods: Heartbeat 2.78 time steps ($\delta_h = 1/2.78$); respiration 18.75.

Here's the fit of this model:



Add a component to filter out the odd pattern in slope

Model 1's "signal" fit showed an unexpected pattern, roughly cyclic with period $\sim$117 time steps.

Let's filter it out of the signal by adding a third quasi-cyclic component:

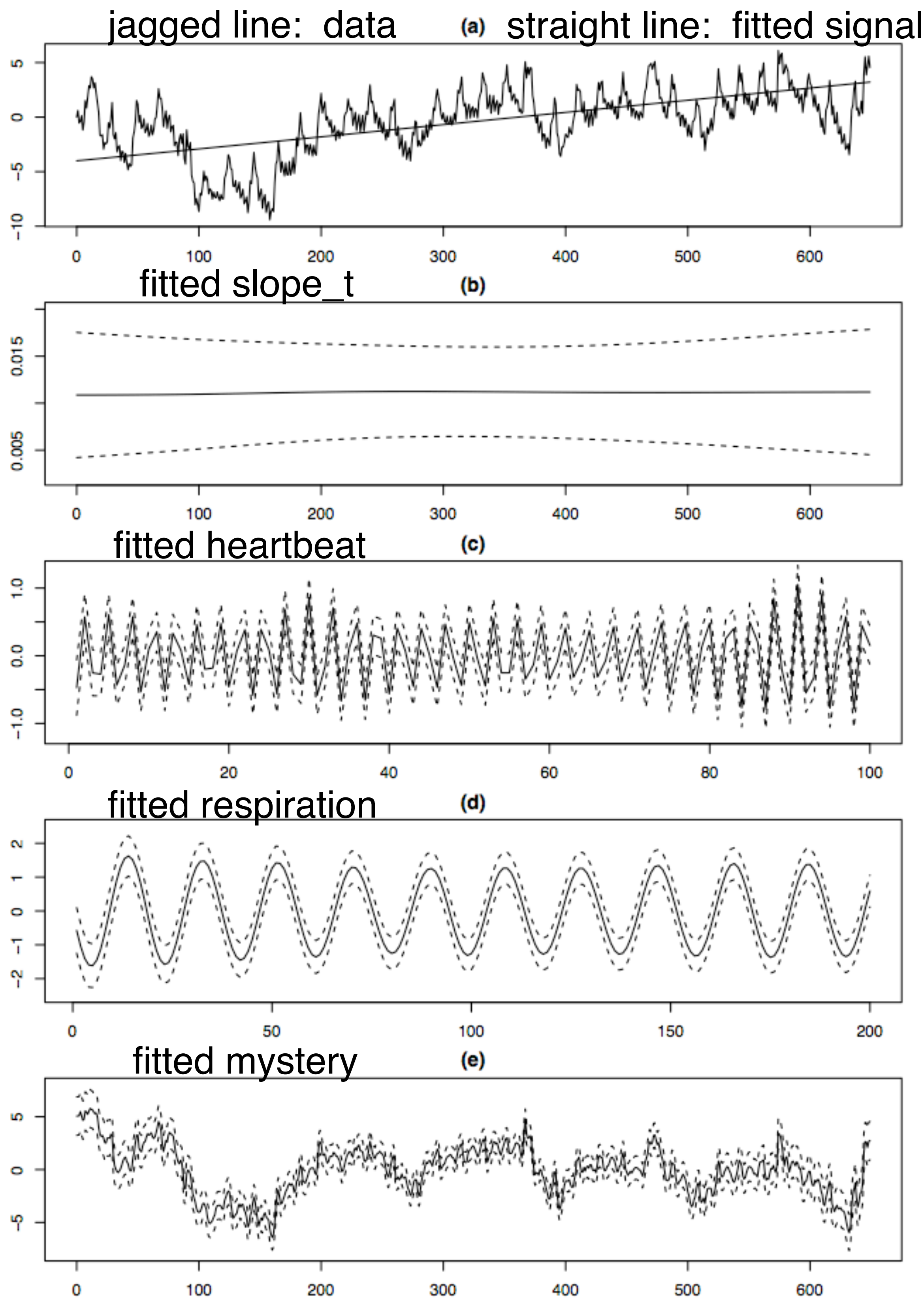
Model 2: $y_t = s_t + h_t + r_t + m_t + v_t$,

where m_t is the new mystery term

The model for m_t has the same form as h_t and r_t with period 117.

Simple, right?

SURPRISE! The mystery term changes everything



Variation formerly captured by signal and respiration are now captured by mystery

What happened? Two possible explanations

(1) The likelihood is bi-modal; the fit really didn't change that much, the fitter just found a different mode.

This appears not to be the case.

(2) The model is spectacularly overparameterized; it's collinearity.

Model 2: $y_t = s_t + h_t + r_t + m_t + v_t$,

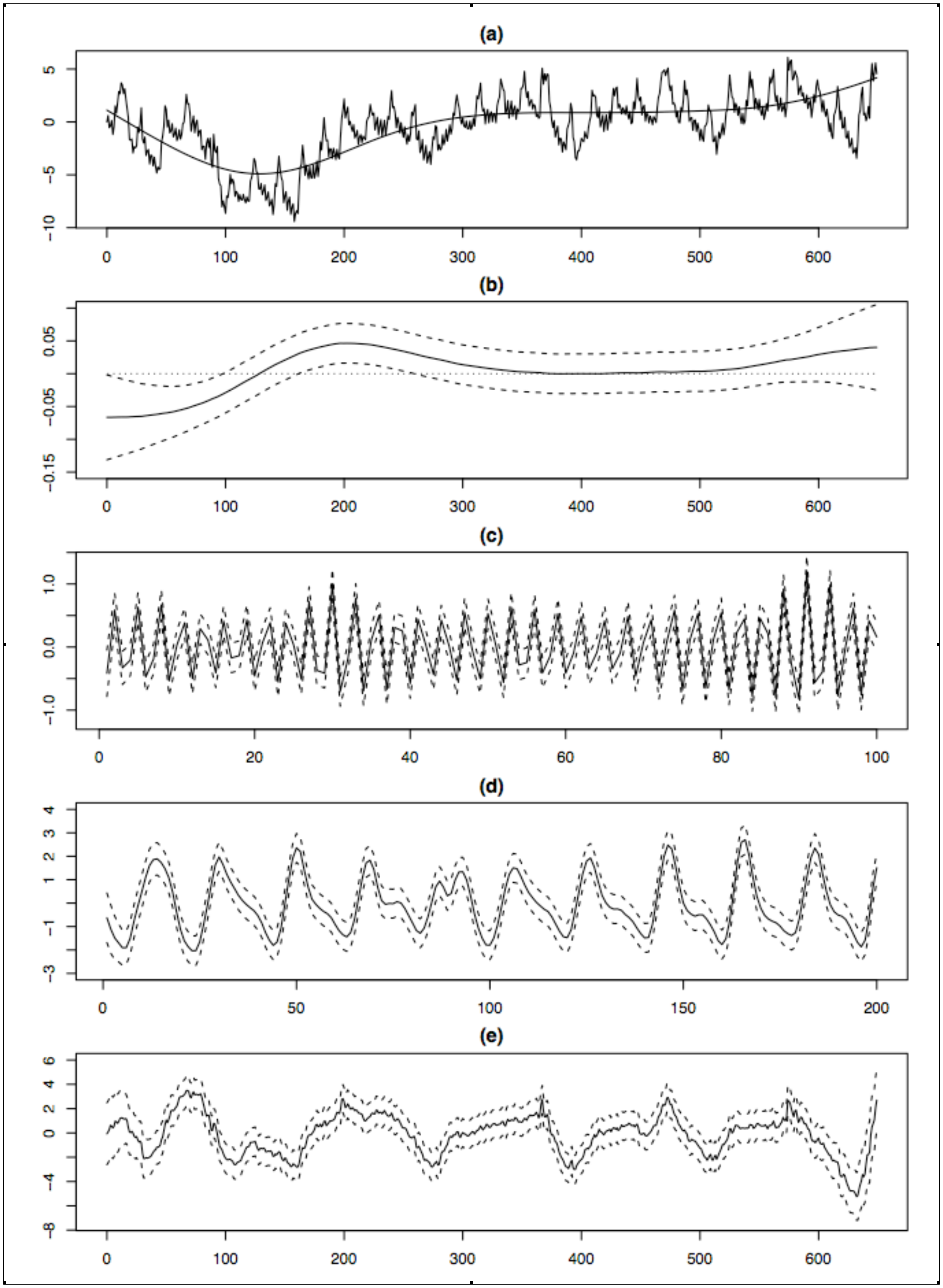
- s_t has n parameters
- h_t, r_t, m_t *each* have $2n$ parameters.

These effects are identified only because they're shrunk/smoothed.

As if all that wasn't weird enough, by inspection the investigators decided to add second harmomics to mystery and respiration ...

Now add second harmonics to mystery and respiration

$$\begin{pmatrix} b_{1,t}\cos\alpha_{1,t} \\ b_{1,t}\sin\alpha_{1,t} \\ b_{2,t}\cos\alpha_{2,t} \\ b_{2,t}\sin\alpha_{2,t} \end{pmatrix} = \begin{bmatrix} \cos\delta & \sin\delta & 0 & 0 \\ -\sin\delta & \cos\delta & 0 & 0 \\ 0 & 0 & \cos 2\delta & \sin 2\delta \\ 0 & 0 & -\sin 2\delta & \cos 2\delta \end{bmatrix} \begin{pmatrix} b_{1,t-1}\cos\alpha_{1,t-1} \\ b_{1,t-1}\sin\alpha_{1,t-1} \\ b_{2,t-1}\cos\alpha_{2,t-1} \\ b_{2,t-1}\sin\alpha_{2,t-1} \end{pmatrix} + \mathbf{w}_{r,t},$$



data & fitted signal

fitted slope_t

fitted heartbeat

fitted respiration

fitted mystery

What's going on? (A road map for this section)

- Detour: Generate hypotheses using a simpler model
- Hypotheses: To get this effect, you need 2 things:
  - Collinearity between the design matrices of the random effects
  - A certain kind of lack of fit
- Collinearity: H2013 Section 12.2.1 shows a sense in which
  - Mystery is more collinear with signal than with respiration
  - All three are more collinear with each other than with heartbeat.
- Lack of Fit
  - I'll use artificial datasets to examine the effects of specific kinds of lack of fit.
- Later, a different tool will let us see lack of fit more clearly

Detour: A model with clustering and heterogeneity

- Model one person's attachment loss measurements
  - 7 teeth, 6 sites per tooth, each site measured twice
  - $7 \times 6 \times 2 = 84$ total measurements

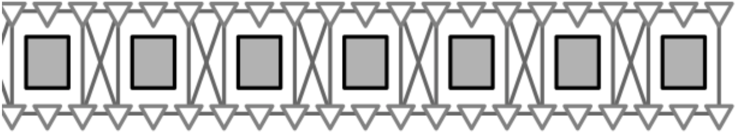
- y_{ik} is the k^{th} measurement of site i, $i = 1, \ldots, N$ for $N = 42$

$$y_{ik} = \delta_i + \xi_i + \epsilon_{ik},$$

- $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_{42})'$ captures spatial clustering
- $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_{42})'$ captures heterogeneity.

- Like the DLMs: Each component has one unknowns for each site.
- Unlike the DLMs: Only two components (clustering + hetero).

- Clustering component: $\boldsymbol{\delta} \sim$ normal ICAR model, $N \times N$ precision matrix $\mathbf{Q}/\sigma_c^2$ and these neighbor pairings:



- Heterogeneity component: $\boldsymbol{\xi} \sim N(\mathbf{0}_N, \sigma_h^2 \mathbf{I}_N)$.
- Sort the observations: 42 first measurements on each site then 42 second measurements on each site.
- Write the model as

$$\mathbf{y} = \begin{bmatrix} \mathbf{I}_N \\ \mathbf{I}_N \end{bmatrix} \boldsymbol{\delta} + \begin{bmatrix} \mathbf{I}_N \\ \mathbf{I}_N \end{bmatrix} \boldsymbol{\xi} + \boldsymbol{\epsilon}. \tag{1}$$

- Clustering and heterogeneity have identical design matrices but different covariance matrices.

Now derive the restricted likelihood

- Let $\mathbf{Q} = \mathbf{\Gamma} \mathbf{D} \mathbf{\Gamma}'$ (spectral decomposition)
  - $\mathbf{D}$ is diagonal, $d_1 \geq d_2 \geq \cdots \geq d_{N-1}$ and $d_N = 0$
  - $\mathbf{\Gamma}$ is orthogonal, N^{th} column (zero eigenvalue) $\propto \mathbf{1}_N$.
- Reparameterize:
  - <u>Clustering</u>: $\boldsymbol{\theta} = \mathbf{\Gamma}'\boldsymbol{\delta}$; $\boldsymbol{\theta}$ has precision $\mathbf{D}/\sigma_c^2$
    $\theta_N = N^{-0.5}\mathbf{1}_N'\boldsymbol{\delta}$ has precision zero.
  - <u>Hetero</u>: $\boldsymbol{\phi} = \mathbf{\Gamma}'\boldsymbol{\xi}$, $\boldsymbol{\phi} \sim N(\mathbf{0}_N, \sigma_h^2 \mathbf{I}_N)$
    ϕ_N is redundant with θ_N, so drop it.
- Now the model is

$$\mathbf{y} = \mathbf{1}_{2N}\beta_0 + \begin{bmatrix} \mathbf{\Gamma}_- \\ \mathbf{\Gamma}_- \end{bmatrix} \boldsymbol{\theta}_- + \begin{bmatrix} \mathbf{\Gamma}_- \\ \mathbf{\Gamma}_- \end{bmatrix} \boldsymbol{\phi}_- + \boldsymbol{\epsilon},$$

- $\mathbf{\Gamma}_-$ is $N \times (N-1)$, the $N-1$ columns of $\mathbf{\Gamma}$ with $d_j > 0$.
- $\boldsymbol{\theta}_-$, $\boldsymbol{\phi}_-$ are $(N-1) \times 1$; they're $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ without N^{th} element.

The restricted likelihood is the likelihood from $\mathbf{\Gamma}'_{-}\mathbf{y}$,

$$\begin{aligned}
\propto \quad & (\sigma_e^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_e^2} SSE\right] && \text{info from replicate measures} \\
& \prod_{j=1}^{N-1} \left(\sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2\right)^{-1/2} && \text{info from } \hat{\theta}_j^2 \\
& \exp\left[-\frac{1}{2}\sum_{j=1}^{N-1} \hat{\theta}_j^2 \left(\sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2\right)^{-1}\right], && \text{info from } \hat{\theta}_j^2
\end{aligned}$$

where

- $SSE = \sum_{i,k}(y_{ik} - \bar{y}_{i.})^2$
- $\bar{y}_{i.}$ is the average of y_{i1} and y_{i2}
- $\hat{\theta}_j$ is the j^{th} element of the unshrunk estimate of $\boldsymbol{\theta}_{-}$

$$\hat{\boldsymbol{\theta}}_{-} = \left[\ \mathbf{\Gamma}'_{-} \quad \mathbf{\Gamma}'_{-}\ \right]\mathbf{y}/2,$$

$$
\begin{aligned}
RL \quad \propto \quad & (\sigma_e^2)^{-N/2} \exp\left[-\frac{1}{2\sigma_e^2} SSE\right] && \text{info from replicate measures} \\
& \prod_{j=1}^{N-1} \left(\sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2\right)^{-1/2} && \text{info from } \hat{\theta}_j^2 \\
& \exp\left[-\frac{1}{2} \sum_{j=1}^{N-1} \hat{\theta}_j^2 \left(\sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2\right)^{-1}\right], && \text{info from } \hat{\theta}_j^2
\end{aligned}
$$

- Note that $\text{var}(\hat{\theta}_j | \sigma_c^2, \sigma_h^2, \sigma_e^2) = \sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2$
- The contributions from heterogeneity and error are the same for all j.
- The contribution from clustering, σ_c^2/d_j, is
  - large for j with small d_j (little smoothing), and
  - small for j with large d_j (much smoothing).

Collinearity + specific lack of fit ⇒ DLM effect

In the het + clust model: $\text{var}(\hat{\theta}_j | \sigma_c^2, \sigma_h^2, \sigma_e^2) = \sigma_c^2/d_j + \sigma_h^2 + \sigma_e^2/2$

Suppose the data actually fit the het + clust model, i.e., the magnitudes of the $\hat{\theta}_j$ decline as the d_j increase.

- If you fit a heterogeneity-only model, $\hat{\sigma}_h^2$ will be big and so will the fitted heterogeneity component.
- If you then fit the het + clust model, $\hat{\sigma}_h^2$ and the fitted heterogeneity component will be small; $\hat{\sigma}_c^2$ and the fitted clustering component will be large.
- Adding clust here is analogous to adding mystery to the DLM.

Thus, a hypothesis: To get the effect we see with the DLM, we need

- The two effects (hetero and clust) are collinear
- The hetero-only model doesn't fit but hetero + clust does.

Back to DLMs: The collinearity part of the hypothesis

H2013, Section 12.2.1 shows a specific sense in which

- Mystery is more collinear with signal than with respiration
- Signal, mystery, and respiration are more collinear with each other than with heartbeat.

Arm-waving explanation:

- Signal, mystery, respiration, and heartbeat have nominal frequencies 0.0015, 0.0085, 0.053, 0.36 per time step.
- Signal and mystery are more similar than any other pair [etc.].

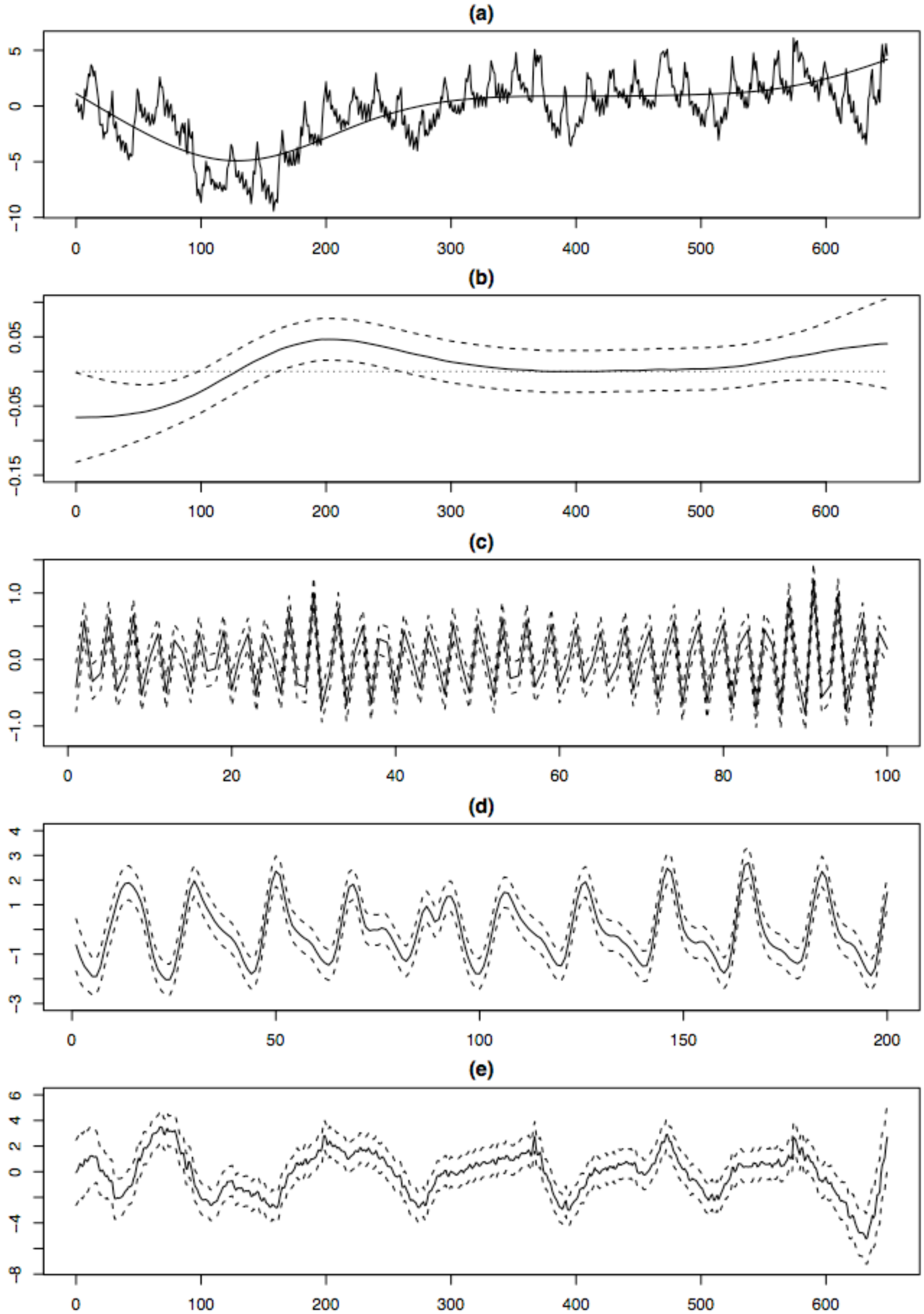
I won't belabor this further; read the book if you're keen to know.

DLMs: Lack of fit (H2013, Section 12.2.2)

- I used Model 3's fit to construct fake datasets with known lack of fit, then fit Models 1 and 2 to see which aspects of lack of fit produced the DLM puzzle.
- Model 3's fit has three pieces that do not fit Model 1:
  - Respiration's 2nd harmonic.
  - Mystery's 1st and 2nd harmonics.
- I constructed fake datasets by adding:
  - Model 3's fitted signal + heartbeat + respiration 1st harmonic
  - + some pieces that do not fit Model 1.

Fake Dataset	respiration 2^{nd} harmonic	mystery 1^{st} harmonic	mystery 2^{nd} harmonic
A	X		
B		X	
C		X	X
Real data	X	X	X

Recap: Here's the fit of Model 3



Data & fitted signal

fitted slope_t (6.5 DF in signal fit)

fitted hb (291 DF)

fitted resp, 2 harmonics (189 DF)

fitted mystery, 2 harmonics (235 DF)

Fake Dataset	respiration 2^{nd} harmonic	mystery 1^{st} harmonic	mystery 2^{nd} harmonic
A	X		
B		X	
C		X	X
Real data	X	X	X

- Models considered
  - Model 1: signal + respiration + heartbeat
  - Model 2: signal + respiration + heartbeat + mystery

- Results for fake dataset A:
  - Model 1: Signal fit smooth, 13 DF; hb and resp like real data;
  - Model 2: Almost unchanged – mystery gets 2.06 DF.

- Results for fake dataset B:
  - Model 1: Signal 42 DF (real data 33), hb favored over resp;
  - Model 2: Signal 17 DF, mystery 39 DF, hb & resp $\sim$unchanged.

Fake Dataset	respiration 2^{nd} harmonic	mystery 1^{st} harmonic	mystery 2^{nd} harmonic
C		X	X
Real data	X	X	X

- Model 1: signal + respiration + heartbeat
- Model 2: signal + respiration + heartbeat + mystery

- Fake dataset C, DF in components of fit

Model	obj fcn	signal	hb	resp	mystery	error
1	49.2	62.6	217.3	137.3	–	232.8
1	60.3	48.6	297.2	304.1	–	0.0005
2		6.6	248.6	56.3	338.5	0.002

This partly reproduces the effects in the real data:

- Respiration is radically smoothed in Model 2.
- Signal is much smoother in Model 2 but not flat as in the real data.

Conclusion re lack of fit

- To reproduce fully the DLM puzzle, lack of fit must include *both*
  - Respiration's 2nd harmonic.
  - Mystery's 1st and 2nd harmonics.
- It's not enough to include only:
  - 2nd harmonic of respiration (Fake data A).
  - 1st harmonic of mystery (Fake data B).
  - 1st and 2nd harmonic of mystery (Fake data C), though this gets closest to reproducing the full effect.