Collinearity/Confounding in richly-parameterized models

For MLM, it often makes sense to consider

$$\mathbf{X}\beta + \mathbf{Zu}$$

to be the mean structure, especially for new-style random effects.

This is like an ordinary linear model, but with $\mathbf{u}$ shrunk toward zero.

The idea of collinearity/confounding from ordinary linear models should be applicable here.

The novelty is

- collinearity of columns in $\mathbf{X}$ (fixed effects) and $\mathbf{Z}$ (random effects);
- $\mathbf{u}$ is shrunk toward zero, to a degree determined as part of the fit.

We'll use collinearity to examine four odd things that happened in real problems.

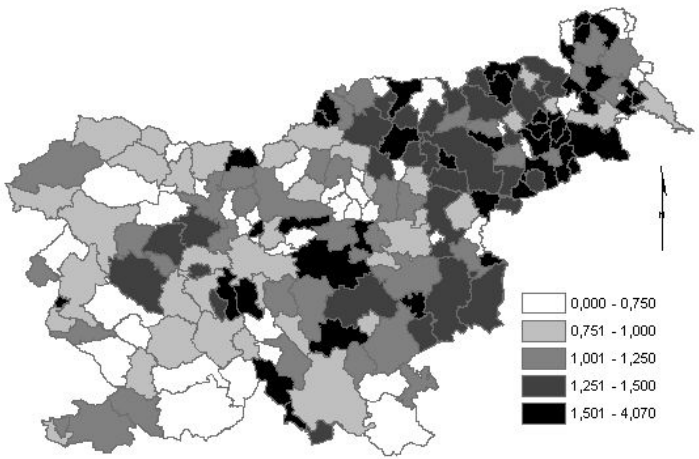
Oddity #1: Add a spatial RE, wipe out a clear association

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

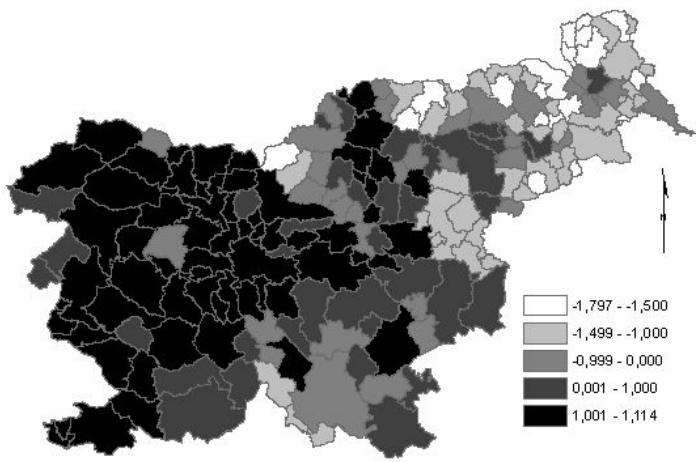
Dataset: For the $i = 1, \ldots, 194$ municipalities that partition Slovenia

- 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.)

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

This model 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 increases, which we expected, and
- β's posterior median to move to 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.

Oddity #2: Adding a random effect changes one fixed effect but not another

The study (kids'n'crowns):

- Badly decayed primary teeth are often capped with a crown.
- Do crown types differ in failure behavior?
- Compare types I, III, IV by time to failure.

The dataset:

- 202 children from pediatric dental practices.
- Each child has between 1 and 4 crowns in the dataset.
- A given child's crowns are all the same type.
- We have covariates (e.g., age) but they don't matter for the present purpose.

Analyses using Cox regression with a random effect

We did analyses both without (wrong) and with (right) an RE for child.

Parameterization: Indicators for Types III and IV (reference is Type I).

Crown Type	Random Effect?	Estimate	Standard Error	P-Value
III	Absent	0.48	0.20	0.015
	Present	0.22	0.41	0.59
IV	Absent	0.14	0.14	0.33
	Present	0.16	0.26	0.54

Estimated SD of child RE is $\sim$1.2; $e^{4.7} = 106 \Rightarrow$ the child effect is big.

Expected: The standard errors got bigger.

Unexpected: One fixed effect estimate changed a lot, the other didn't.

Why?

Oddity #3: Differential shrinkage of equal-sized effects in smoothed ANOVA

Dataset: ∼2900 people with colon cancer, after surgery to remove tumors (combining 7 clinical trials of the same treatment)

Question: We know there's a treatment *main effect*; does the tx effect depend on patient age (4 groups) and cancer stage (II vs. III)?

Analysis:

- Outcome: Disease-free survival (event = progression or death)
- Analysis:
  - Include all interactions and shrink them (smoothed ANOVA).
  - Mostly Bayesian, but using Cox's partial likelihood.
  - Design-matrix columns were scaled (same Euclidean length).

	No shrinkage		Shrinkage	
Effect	Est	Interval	Est	Interval
treatment-by-stage	-4.2	$(-8.3, 0.02)$	-2.5	$(-5.9, 0.3)$
treatment-by-age 1	-4.6	$(-8.8, -0.5)$	-2.9	$(-5.9, 0.3)$
treatment-by-age 2	-4.2	$(-8.3, 0.01)$	-0.6	$(-2.6, 0.6)$
treatment-by-age 3	-4.8	$(-9.0, -0.6)$	-1.1	$(-5.7, 1.8)$
stage main effect	-25.9	$(-30.1, -21.8)$	-23.4	$(-26.7, -20.0)$

"Est" is the posterior mean; "Interval" is an equal-tailed 95% interval.

Unsmoothed CIs are all about the same width.

Why are the effects shrunk to different extents?

- In balanced SANOVA with normal errors, this can't happen.
- But here, design matrix columns are not orthogonal, in two senses:
  - The design is not balanced.
  - The error variance is not independent of the design-cell mean.

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

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

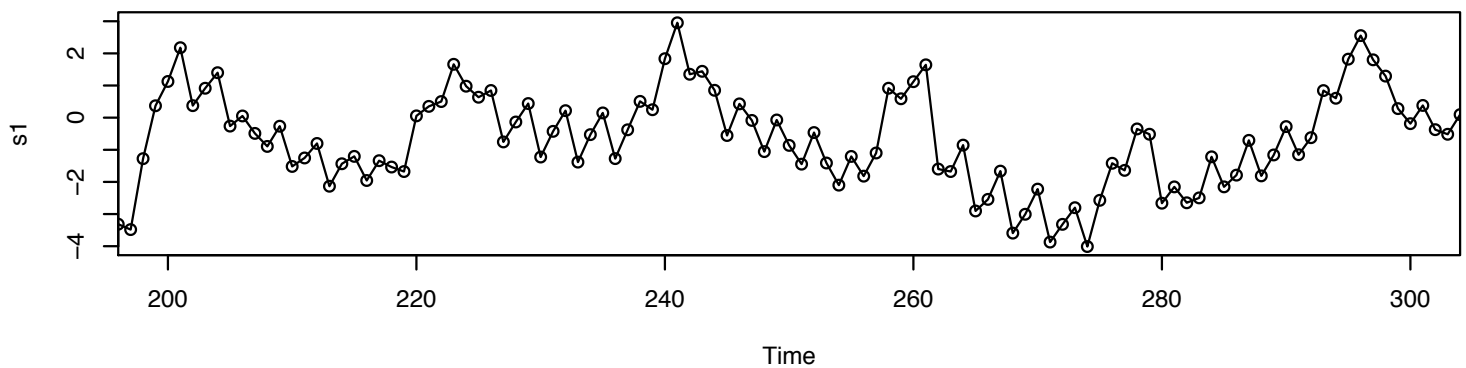
y_t = % change in average pixel value for light of wavelength 535 nm, $t = 0, \ldots, 649$, with time steps of 0.28 sec.

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

Object: Estimate the response to the stimulus.

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

Here's about 100 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 is the smoothed response, the object of this analysis;
- h_t, r_t are heartbeat and respiration respectively;
- v_t is iid $N(0, W_v)$ error.

State equations for s_t, h_t, r_t

State equation for s_t is the linear growth model:

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

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

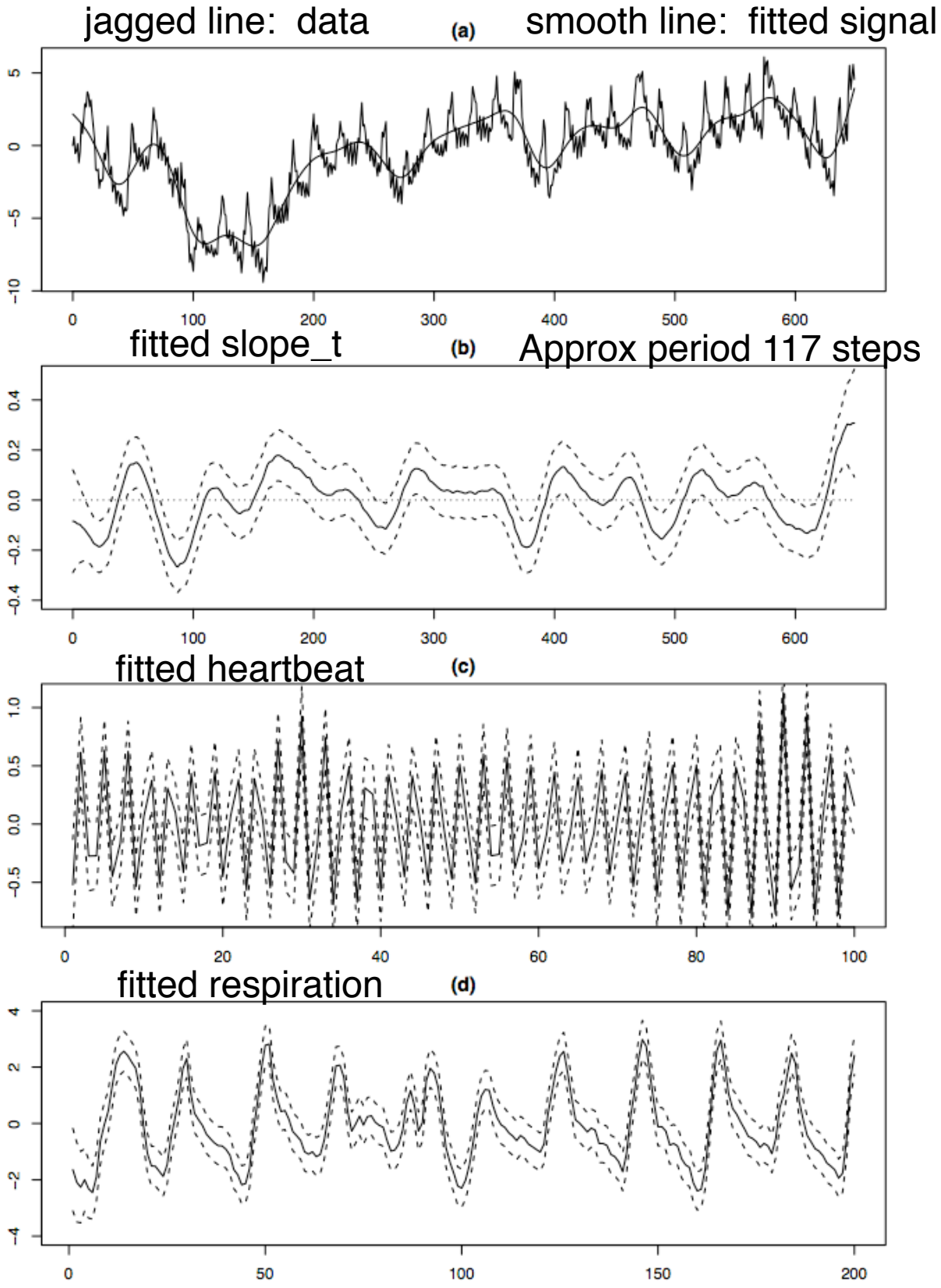
State equation for quasi-cyclic components (this is for heartbeat):

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

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

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:

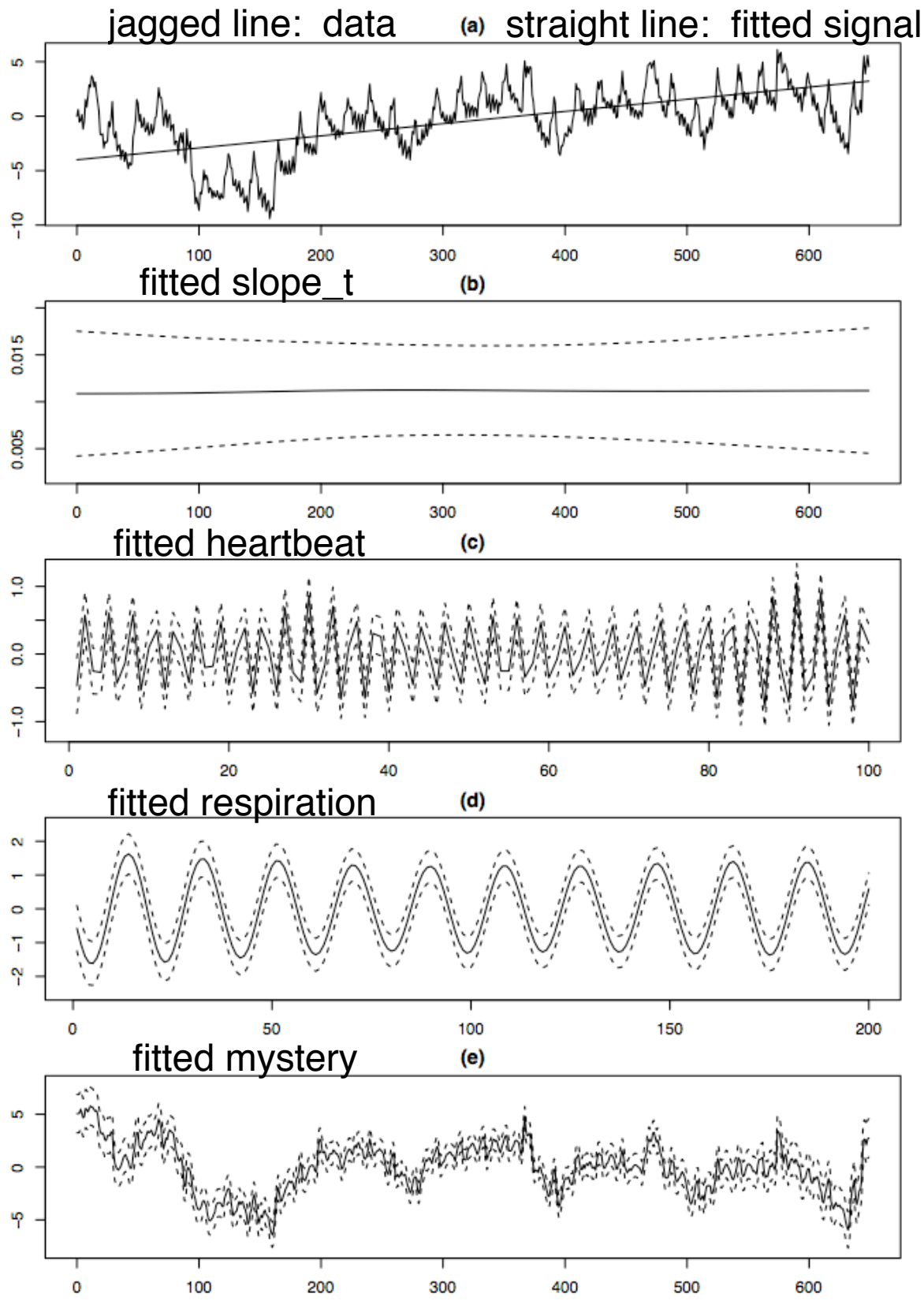
Model 2: $y_t = s_t + h_t + r_t + m_t + v_t$,

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

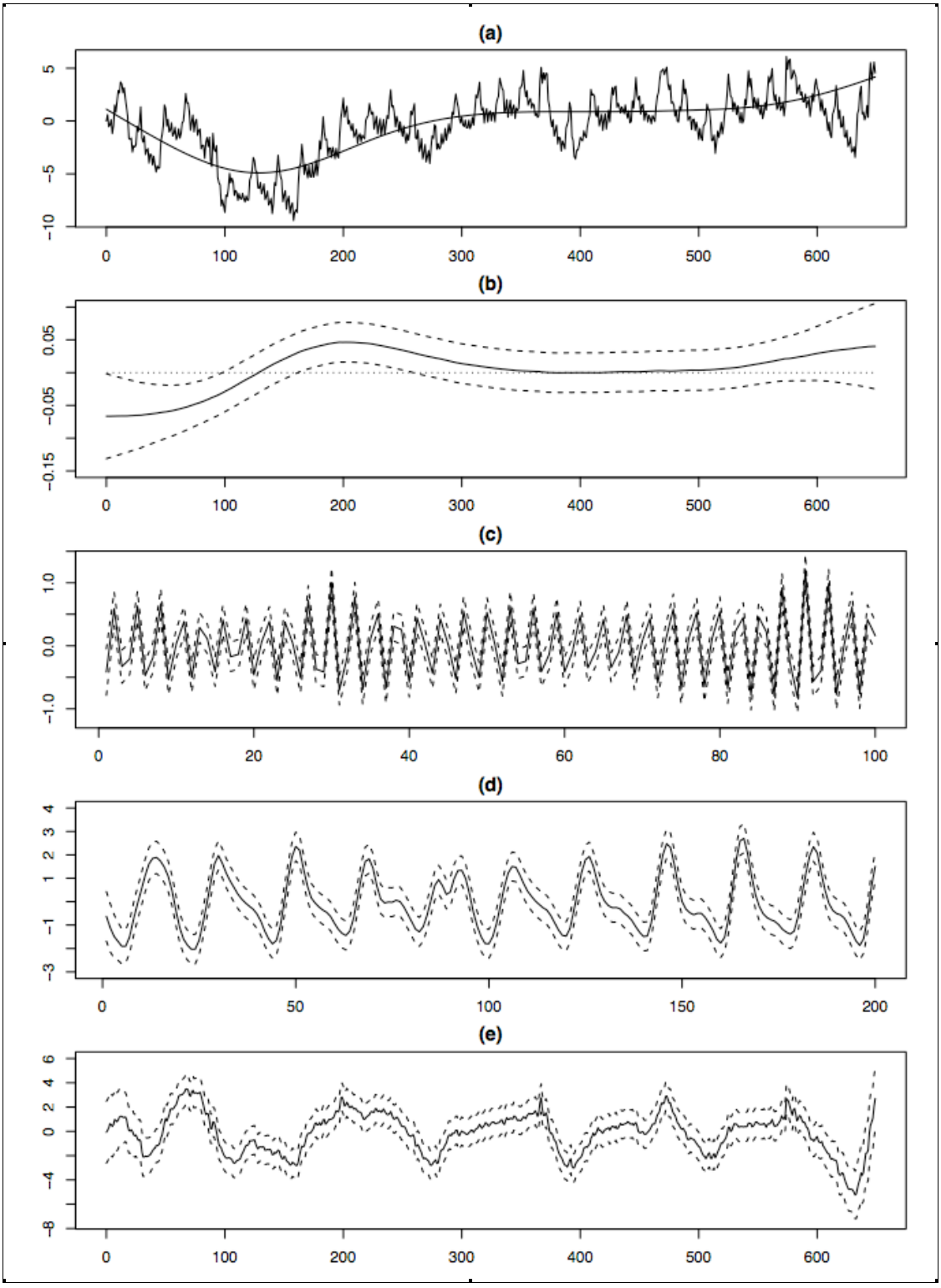
- s_t has n parameters
- h_t, r_t, m_t <u>each</u> have $2n$ parameters.

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

Collinearity/Confounding in richly-parameterized models

For MLM, it often makes sense to consider

$$\mathbf{X}\beta + \mathbf{Zu}$$

to be the mean structure, especially for new-style random effects.

This is like an ordinary linear model, but with $\mathbf{u}$ shrunk toward zero.

The idea of collinearity/confounding from ordinary linear models should be applicable here.

The novelty is

- collinearity of columns in $\mathbf{X}$ (fixed effects) and $\mathbf{Z}$ (random effects);
- $\mathbf{u}$ is shrunk toward zero, to a degree determined as part of the fit.

We'll use collinearity to examine four odd things that happened in real problems.