# Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake*        Yoonkyung Lee†

March 10, 2018

## 1   Cattle weights

See Kenward, Michael G. (1987). A Method for Comparing Profiles of Repeated Measurements. Applied Statistics, 36, 296-308. Table 1. http://doi.org/10.2307/2347788

Kenward [1987] reported an experiment designed to investigate the impact of the control of intestinal parasites in cattle. During the grazing season which runs from spring to autumn, cattle can potentially ingest roundworm larvae, developed from eggs deposited around the pasture from feces of previously infected cattle. Immune resistance to disease is suppressed once an animal is infected and deprived of nutrients; consequently, animal growth can be greatly impacted. Monitoring the effect of a treatment for the disease requires repeated weight measurements on animals over the grazing season.

To compare two methods for controlling the disease, say treatment A and treatment B, each of 60 cattle were assigned randomly to two groups, each of size 30. Animal subjects were put out to pasture at the start of grazing season, with each member of the groups receiving one of the two treatments. Animals were weighed $m = 11$ times over a 133-day period; the first 10 measurements on each animal were made at two-week intervals and the final measurement was made one week later. Weights were recorded to the nearest kilogram. The measurement times were common across animals and were rescaled to $t = 1, 2, \ldots, 10, 10.5$. The longitudinal dataset is balanced, as there were no missing observations for any of the experimental units. Observed weights are shown in Figure **??**.

We see an upward trend in weights over time, with variance in weights increasing over time for both groups. Treatment group B demonstrates a sharp decrease in the final weight measurement. The analysis of the same dataset provided by Zimmerman and Núñez-Antón [1997] rejected

---

*The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201
†The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

Figure 1: Subject-specific weight curves over time for treatment groups A and B.
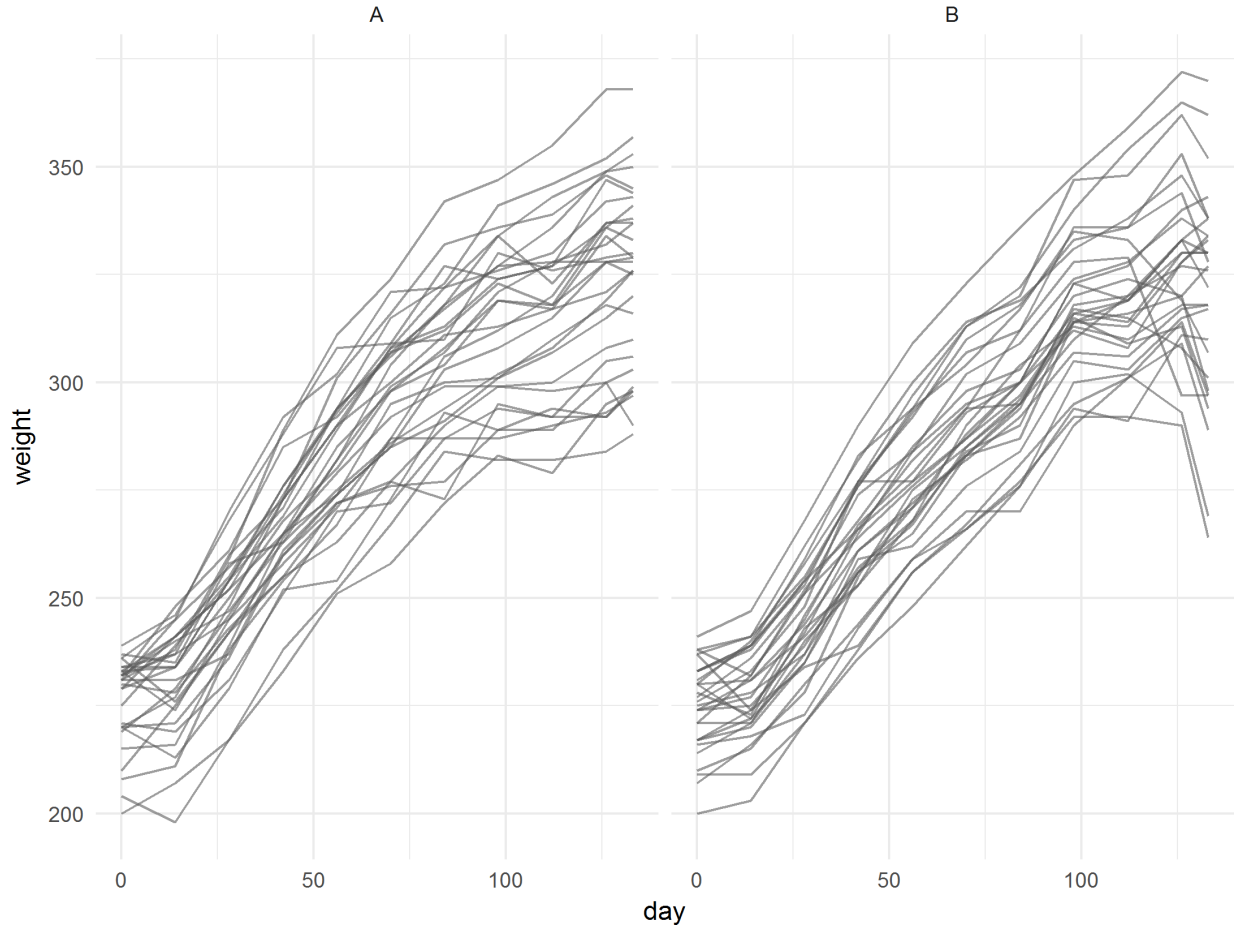


equality of the two covariance matrices corresponding to treatment group using the classical likelihood ratio test, making it reasonable to study each treatment group's covariance matrix separately. Following Pan and Pan [2017], Zhang et al. [2015], Pourahmadi [1999], and Pan and [2006], we analyze the data from the $N = 30$ cattle assigned to treatment group A, which we assume share a common $11 \times 11$ covariance matrix $\Sigma$. The profile plot in Figure **??** of the weights for units in treatment group A shows a clear upward trend in weights; variances appear to increase over time, suggesting that the covariance structure is nonstationary.

Before modeling the covariance structure, it is necessary to construct an adequate estimate of the mean weight trajectories. After centering the data using the fitted mean, the residuals serve as the data reserved for estimating the functions defining the Cholesky factor and innovation variances. To account for any between-subject variability, we adopt an approach akin to the dynamical conditionally linear mixed model proposed by Pourahmadi and Daniels; see Pourahmadi and

Daniels [2002] for a detailed discussion. We model

$$r\left(t_{ij}\right) = \sum_{k<j} \phi\left(t_{ij}, t_{ik}\right) r\left(t_{ij}\right) + \epsilon\left(t_{ij}\right) \tag{1}$$

where

$$r\left(t_{ij}\right) = y\left(t_{ij}\right) - \left(f\left(t_{ij}\right) + \alpha_i\right). \tag{2}$$

The subject-specific random effects $\{\alpha_i\}$ are assumed to be mutually independent and independent of $\epsilon\left(t_{ij}\right)$ for all $i,\ j$, with

$$\alpha_i \sim N\left(0, \sigma_\alpha^2\right).$$

We assume that the subject trajectories share a common mean function $f \in$

$$\mathcal{C}^2 = \left\{ f: \ f,\ f' \text{ absolutely continuous}, \ \int \left(f''\left(x\right)\right)^2 \ dx < \infty \right\}.$$

Figure **??** displays the he observed weight trajectories over time. Figure **??** shows the corresponding fitted mean curves.
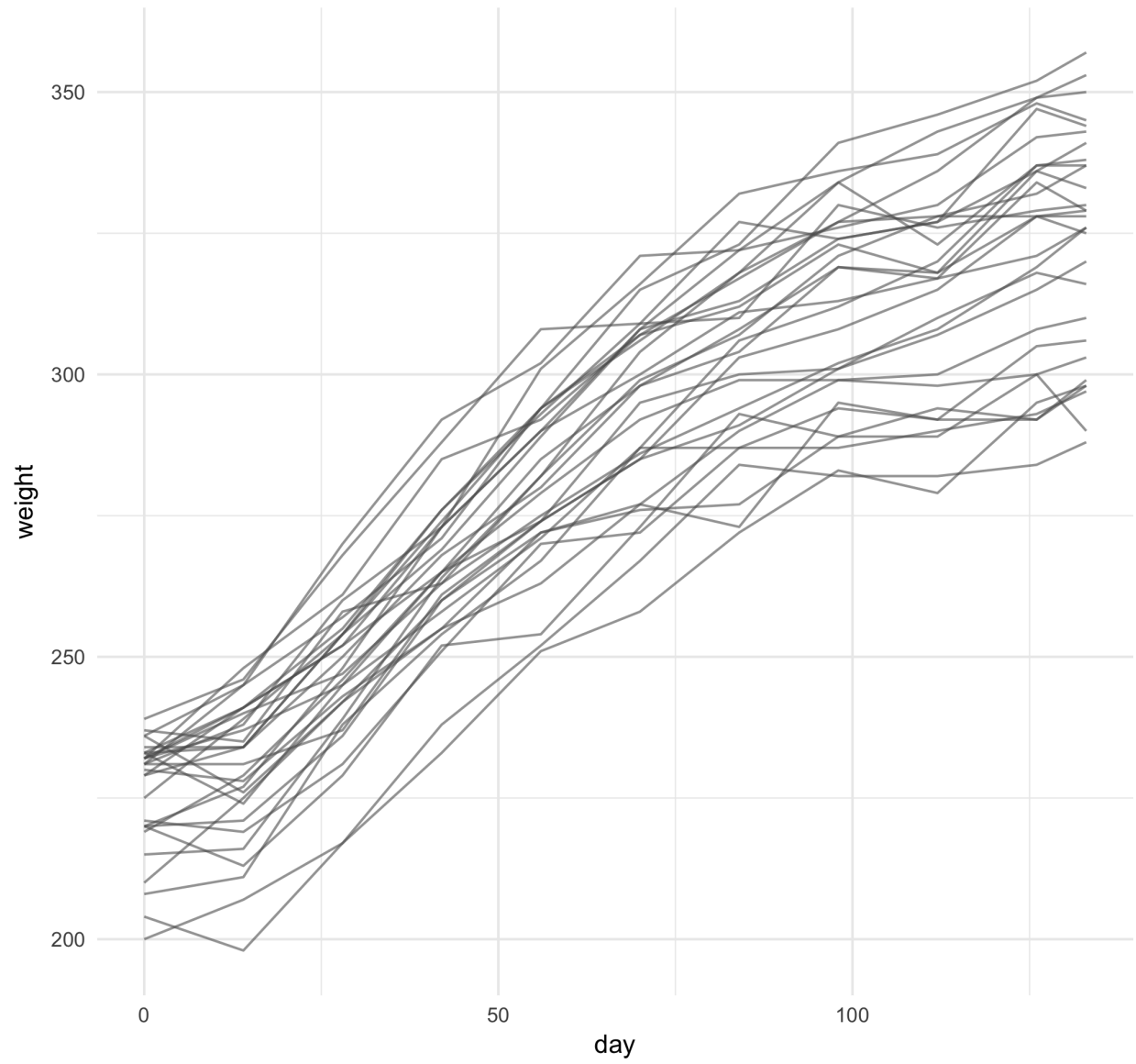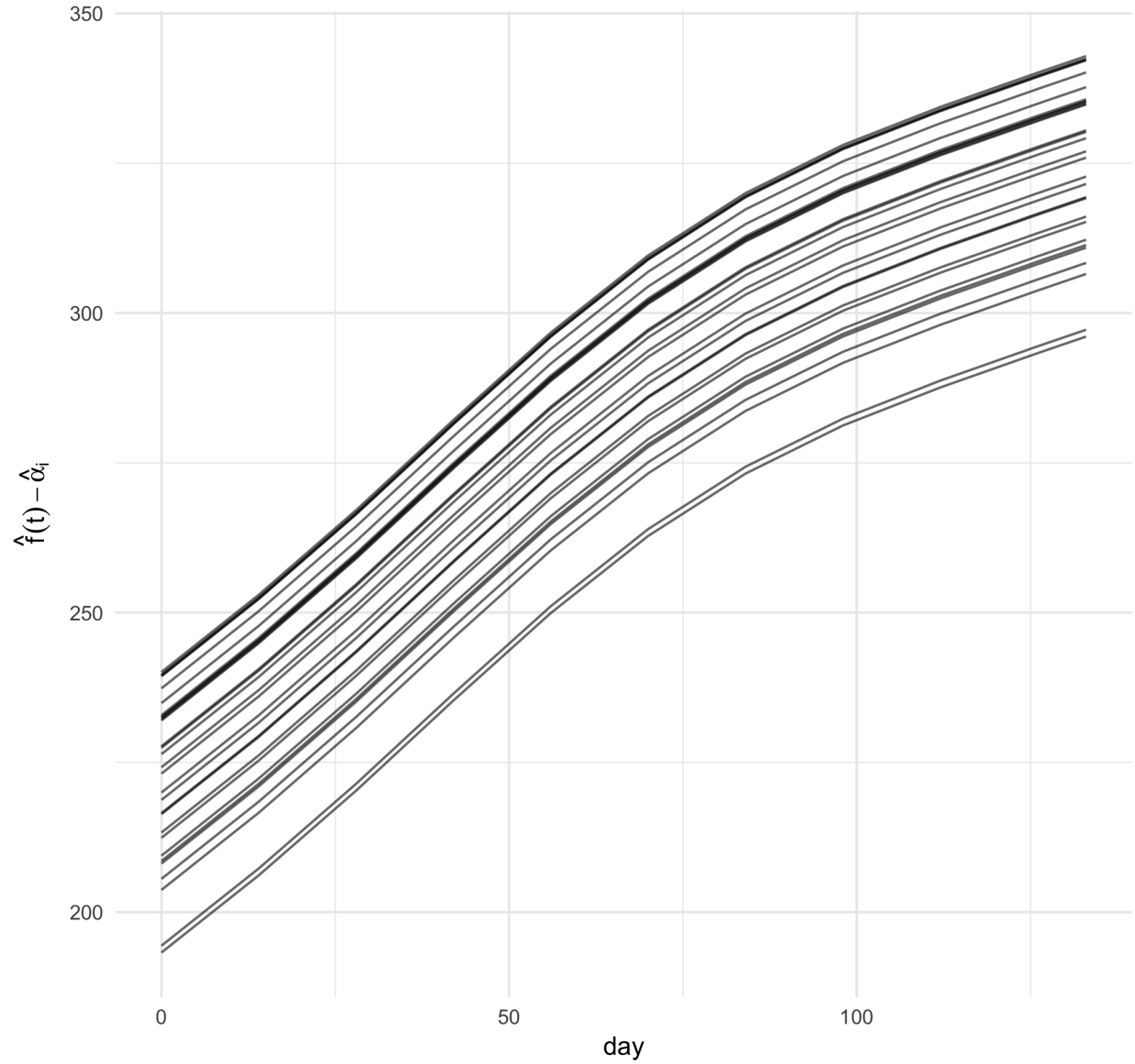
Figure 2: Weight trajectories over the observation period for experimental units in treatment group A.

Figure 3: Fitted mean weight curve for cattle in treatment group A.



The nonstationarity suggested in Figure **??** is also supported by the sample correlations given in Table **??**; correlations within the subdiagonals are not constant and increase over time, a secondary indication that a stationary covariance is not appropriate for the data. Table **??** gives the sample generalised autoregressive parameters and the innovation variances, which are plotted in Figure **??** and Figure **??** respectively.

|     | day | | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|     | 0 | 14 | 28 | 42 | 56 | 70 | 84 | 98 | 112 | 126 | 133 |
| 0 | 1.00 | 0.82 | 0.76 | 0.65 | 0.63 | 0.58 | 0.51 | 0.52 | 0.51 | 0.46 | 0.46 |
| 14 | 0.82 | 1.00 | 0.91 | 0.86 | 0.83 | 0.75 | 0.64 | 0.68 | 0.61 | 0.59 | 0.56 |
| 28 | 0.76 | 0.91 | 1.00 | 0.93 | 0.89 | 0.85 | 0.75 | 0.77 | 0.71 | 0.69 | 0.67 |
| 42 | 0.65 | 0.86 | 0.93 | 1.00 | 0.93 | 0.90 | 0.80 | 0.82 | 0.74 | 0.70 | 0.67 |
| 56 | 0.63 | 0.83 | 0.89 | 0.93 | 1.00 | 0.94 | 0.85 | 0.88 | 0.81 | 0.77 | 0.74 |
| 70 | 0.58 | 0.75 | 0.85 | 0.90 | 0.94 | 1.00 | 0.92 | 0.93 | 0.89 | 0.85 | 0.81 |
| 84 | 0.51 | 0.64 | 0.75 | 0.80 | 0.85 | 0.92 | 1.00 | 0.92 | 0.92 | 0.86 | 0.84 |
| 98 | 0.52 | 0.68 | 0.77 | 0.82 | 0.88 | 0.93 | 0.92 | 1.00 | 0.96 | 0.94 | 0.91 |
| 112 | 0.51 | 0.61 | 0.71 | 0.74 | 0.81 | 0.89 | 0.92 | 0.96 | 1.00 | 0.96 | 0.95 |
| 120 | 0.46 | 0.59 | 0.69 | 0.70 | 0.77 | 0.85 | 0.86 | 0.94 | 0.96 | 1.00 | 0.98 |
| 133 | 0.46 | 0.56 | 0.67 | 0.67 | 0.74 | 0.81 | 0.84 | 0.91 | 0.95 | 0.98 | 1.00 |

Table 1: Cattle data: treatment group A sample correlations.

|   |   | s | | | | | | | | | | |   |   |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|   |   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 10.5 |   |   |
|   | 1 | 1 |   |   |   |   |   |   |   |   |   |   | 4.673 |   |
|   | 2 | 1.00 | ⋱ |   |   |   |   |   |   |   |   |   | 3.939 |   |
|   | 3 | 0.04 | 0.90 |   |   |   |   |   |   |   |   |   | 3.370 |   |
|   | 4 | -0.25 | 0.25 | 0.88 |   |   |   |   |   |   |   |   | 3.000 |   |
|   | 5 | -0.02 | 0.07 | 0.12 | 0.90 |   |   |   |   |   |   |   | 3.299 |   |
| t | 6 | 0.04 | -0.28 | 0.11 | 0.37 | 0.82 |   |   |   |   |   |   | 3.363 | $\log\left(\sigma_t^2\right)$ |
|   | 7 | 0.12 | -0.23 | 0.04 | -0.16 | 0.08 | 1.03 |   |   |   |   |   | 3.610 |   |
|   | 8 | -0.06 | 0.05 | 0.02 | -0.27 | 0.23 | 0.61 | 0.42 |   |   |   |   | 3.403 |   |
|   | 9 | 0.18 | -0.10 | 0.05 | -0.26 | -0.10 | 0.03 | 0.30 | 0.93 |   |   |   | 2.780 |   |
|   | 10 | -0.26 | 0.15 | 0.45 | -0.33 | -0.19 | 0.01 | -0.18 | 0.37 | 0.94 | ⋱ |   | 3.280 |   |
|   | 10.5 | 0.13 | -0.26 | 0.08 | 0.28 | 0.04 | -0.36 | -0.05 | -0.07 | 0.37 | 0.85 | 1 | 2.262 |   |

Table 2: Cattle data: treatment group A sample generalized autoregressive parameters (below the main diagonal) and log sample innovation variances (rightmost column.)
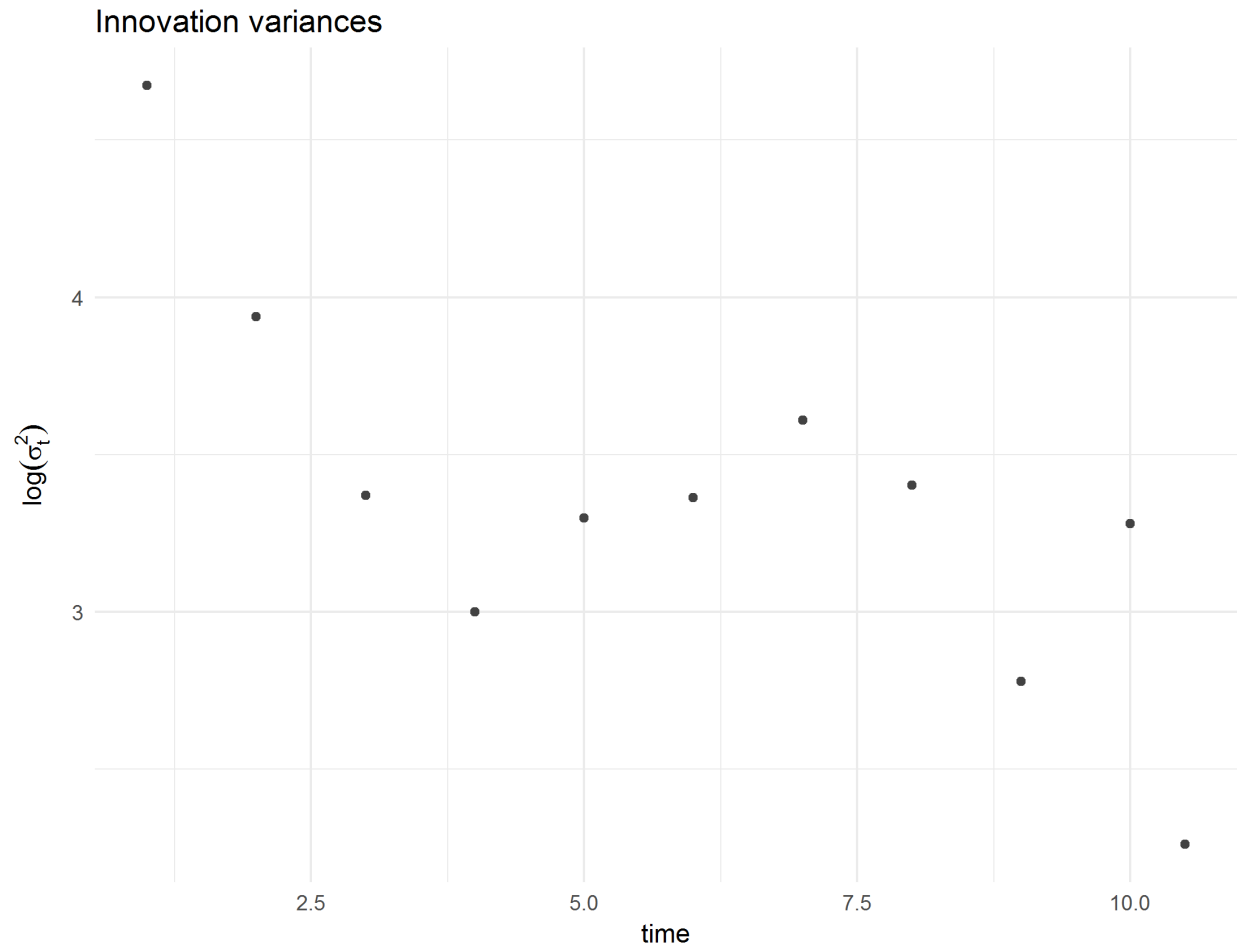
Figure 4: Sample estimates of innovation variances $\sigma_t^2$ obtained by applying the modified Cholesky decomposition to the sample covariance matrix.
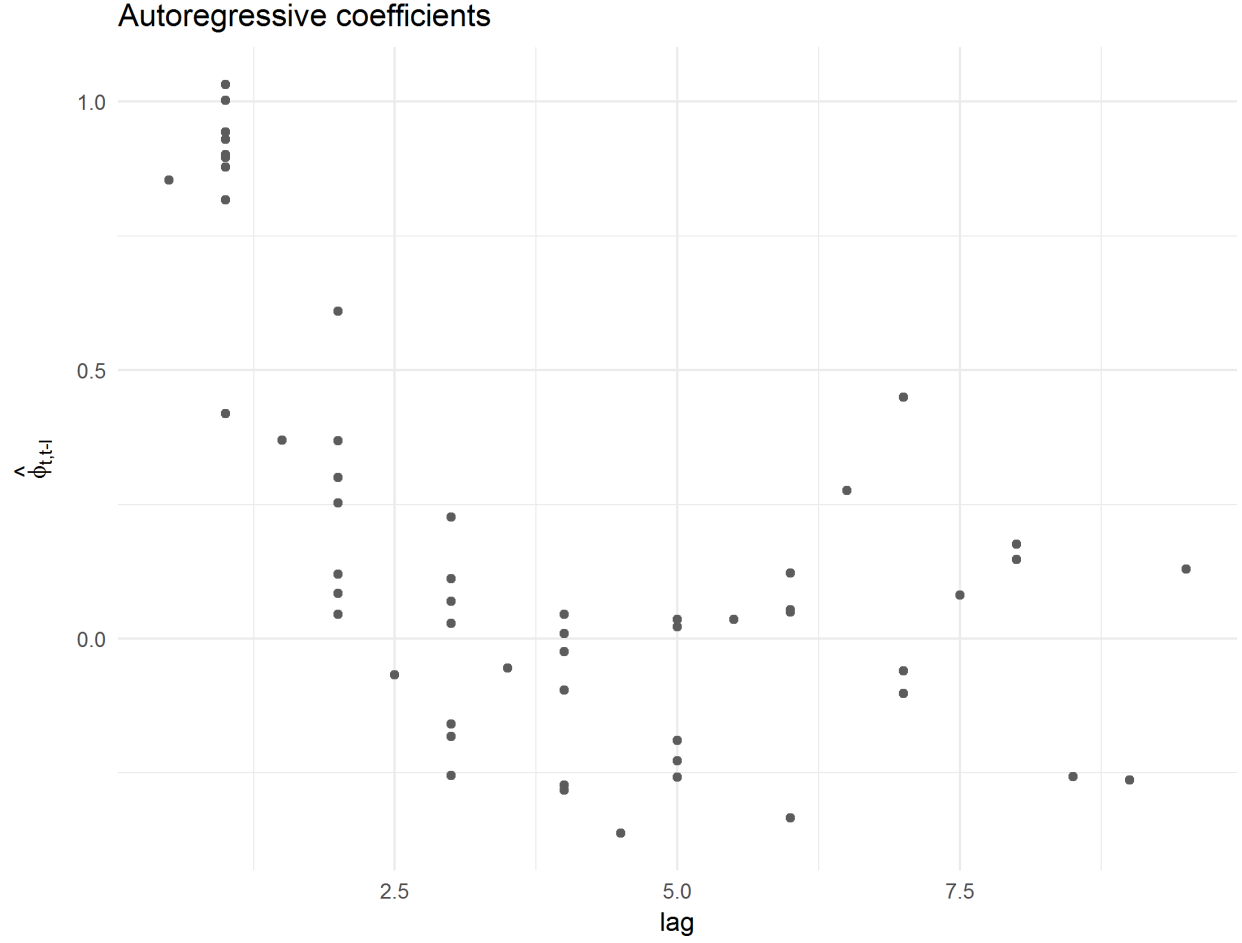
Figure 5: Sample estimates of the generalized autoregressive parameters $\phi_{ij}$ obtained by applying the modified Cholesky decomposition to the sample covariance matrix.

The analysis of Pourahmadi [1999] concluded that the regressogram (Figure **??**) and variogram (Figure **??**) suggest that both sample generalised autoregressive parameters and the logarithms of the innovation variances can be characterized in terms of cubic functions of the lag only. They model

$$
\phi_{ts} = x'_{ts}\gamma,
$$
$$
\log\left(\sigma_t^2\right) = z'_t\xi,
$$

(3)

for $t = 2, \ldots, 11$ where

$$
x'_{ts} = \begin{bmatrix} 1 & t-s & (t-s)^2 & (t-s)^3 \end{bmatrix}, \text{ and } z'_t = \begin{bmatrix} 1 & t & t^2 & t^3 \end{bmatrix}.
$$

8

They estimate of $\gamma$ and $\xi$ via maximum likelihood. Figure **??** shows the estimated cubic polynomials corresponding to Model 3.
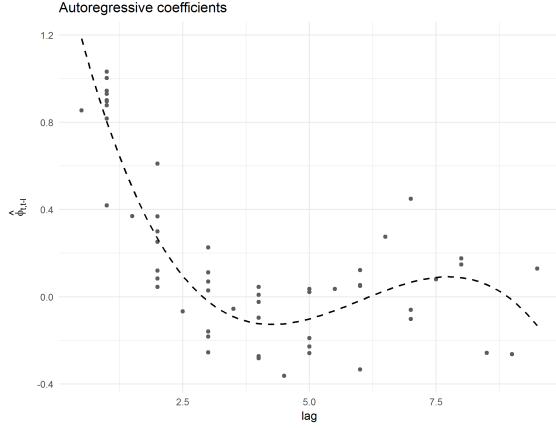


(a) The sample regressogram for the cattle data from treatment group A, overlaid with a cubic polynomial smooth.
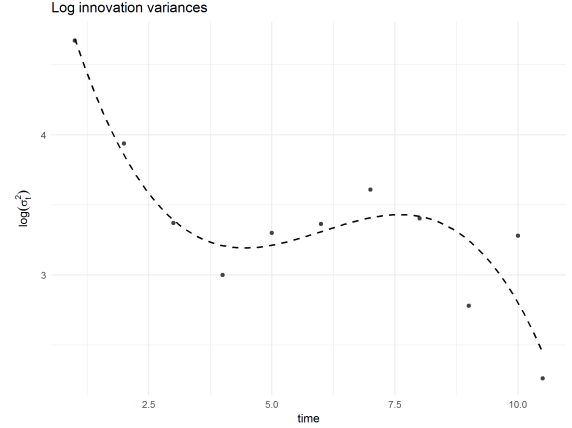
(b) The sample variogram for the cattle data from treatment group A, overlaid with a cubic polynomial smooth.

Choice of penalty is critical for convergence of the iterative estimation of $\phi$ and $\log(\sigma_2)$. Pan and Pan [2017] concluded that the regressogram of empirical estimates of $\phi_{t,t-l}$ show consistent behaviour over $l$ for each value of $t$, indicating a lack of a strong functional component of $m$. To facilitate a making this modeling decision in an entirely data-driven manner, we let $\phi \in \mathcal{H} = \mathcal{H}_{[1]} \otimes \mathcal{H}_{[2]}$, where

$$\mathcal{H}_{[1]} = \left\{ \phi : \ddot{\phi} = 0 \right\} \oplus \left\{ \phi : \phi(0) = \dot{\phi}(0) = 0; \quad \int_0^1 \ddot{\phi}^2 \, dx < \infty \right\}$$

$$\mathcal{H}_{[2]} = \left\{ \phi : \phi \propto 1 \right\} \oplus \left\{ \phi : \int_0^1 \phi \, dx = 0, \quad \dot{f} \in \mathcal{L}_2[0,1] \right\}$$

This decomposition leads to a null space comprised of functions of $l$ only, which is attractive because it coincides with the modeling assumptions made by $\phi$ Pan and Pan [2017], Huang et al. [2006], and Wu and Pourahmadi [2003] for the same data set. Figure 6 and Figure 7 show the estimated Cholesky surface $\phi(t, s)$ and innovation variance function $\sigma^2(t)$ evaluated at $t = 1, 2, \ldots, 10, 10.5$ and the corresponding pairs of observation times $(t, s)$, $1 \leq s < t \leq 10.5$.

Figure 6: Components of the fitted modified Cholesky decomposition for the cattle weight data.
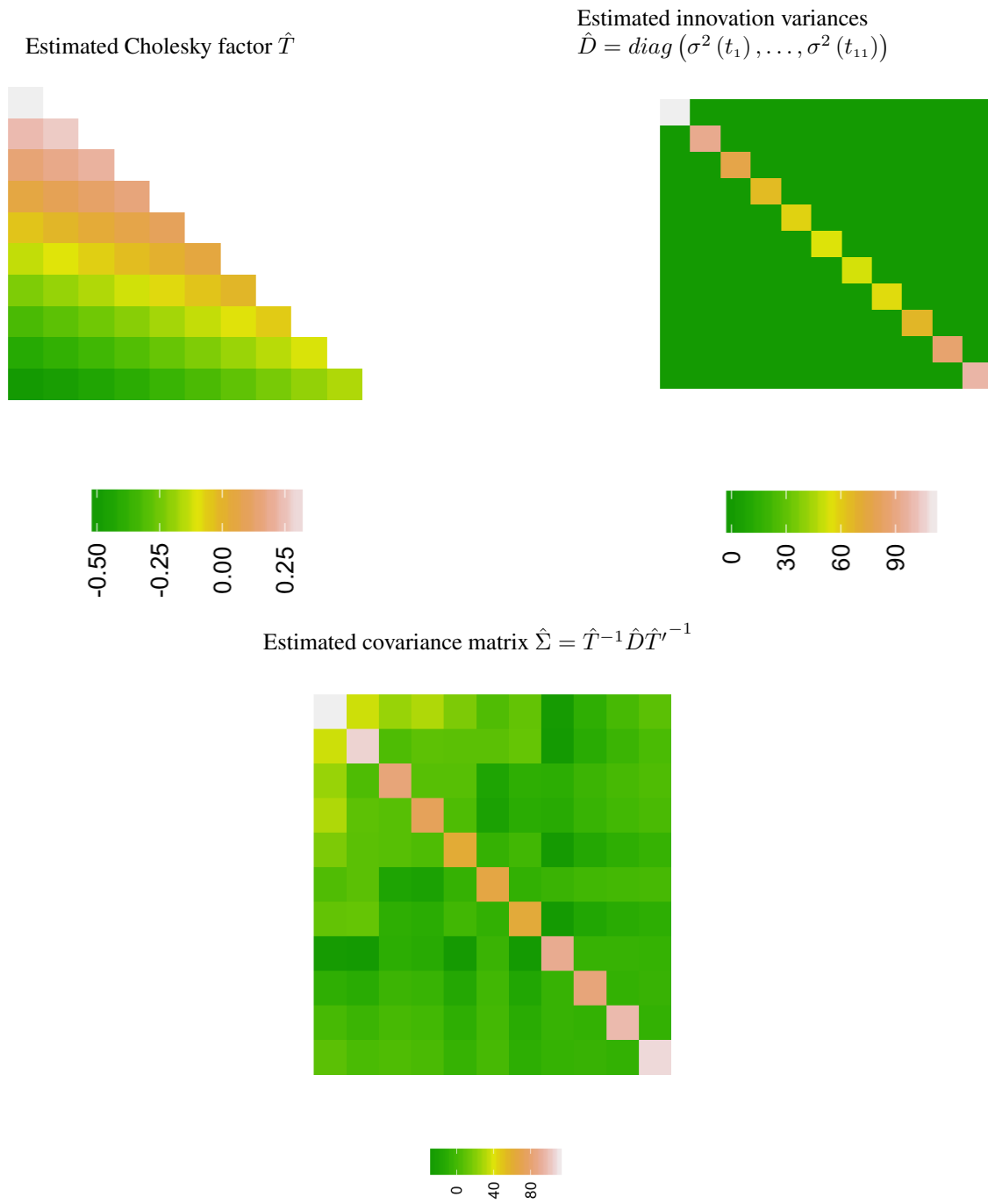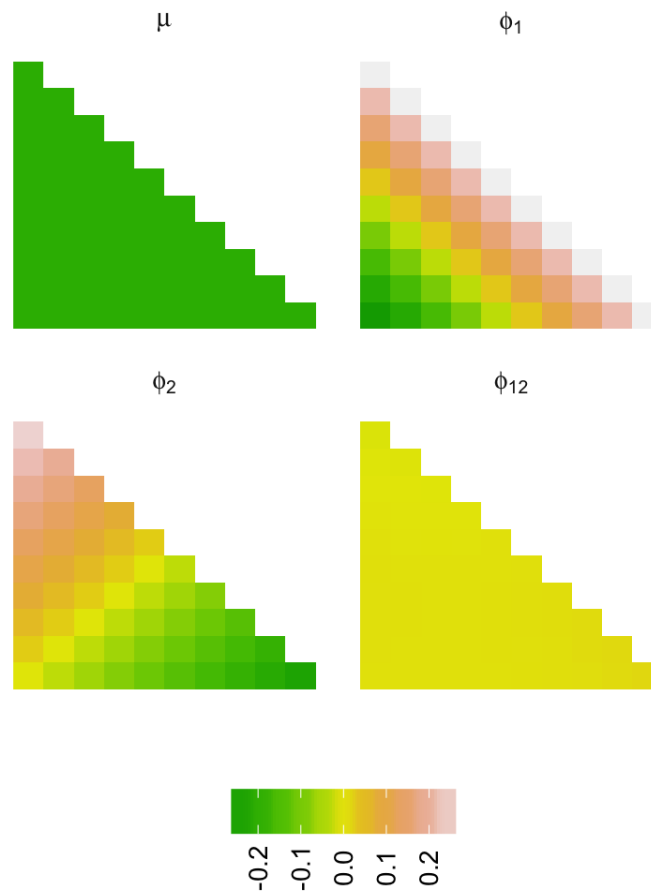
Estimated Cholesky factor $\hat{T}$

Estimated innovation variances
$\hat{D} = diag\left(\sigma^2\left(t_1\right), \ldots, \sigma^2\left(t_{11}\right)\right)$

Estimated covariance matrix $\hat{\Sigma} = \hat{T}^{-1}\hat{D}\hat{T}'^{-1}$

Figure 7: Components of the SSANOVA decomposition of the estimated generalized autoregressive coefficient function $\phi$ evaluated on the grid defined by the observed time points.



# References

Jianhua Z Huang, Naiping Liu, Mohsen Pourahmadi, and Linxu Liu. Covariance matrix selection and estimation via penalised normal likelihood. *Biometrika*, pages 85–98, 2006.

Michael G Kenward. A method for comparing profiles of repeated measurements. *Applied Statistics*, pages 296–308, 1987.

Jianxin Pan and Gilbert . Regression models for covariance structures in longitudinal studies. *Statistical Modelling*, 6(1):43–57, 2006.

Jianxin Pan and Yi Pan. jmcm: An r package for joint mean-covariance modeling of longitudinal data. *Journal of Statistical Software*, 82(1):1–29, 2017.

M Pourahmadi and MJ Daniels. Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58(1):225–231, 2002.

Mohsen Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.

Wei Biao Wu and Mohsen Pourahmadi. Nonparametric estimation of large covariance matrices of longitudinal data. *Biometrika*, 90(4):831–844, 2003.

Weiping Zhang, Chenlei Leng, and Cheng Yong Tang. A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):219–238, 2015.

Dale L Zimmerman and Vicente Núñez-Antón. Structured antedependence models for longitudinal data. In *Modelling longitudinal and spatially correlated data*, pages 63–76. Springer, 1997.