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

Tayler A. Blake*        Yoonkyung Lee†

February 1, 2018

## 1   Cattle weights

Fitted models can be found in /Users/taylerblake/GitRepos/cautious-guacamole/cache/kenward-cattle-analysis.RData

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

? 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 ? rejected equality of the two

---
*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.
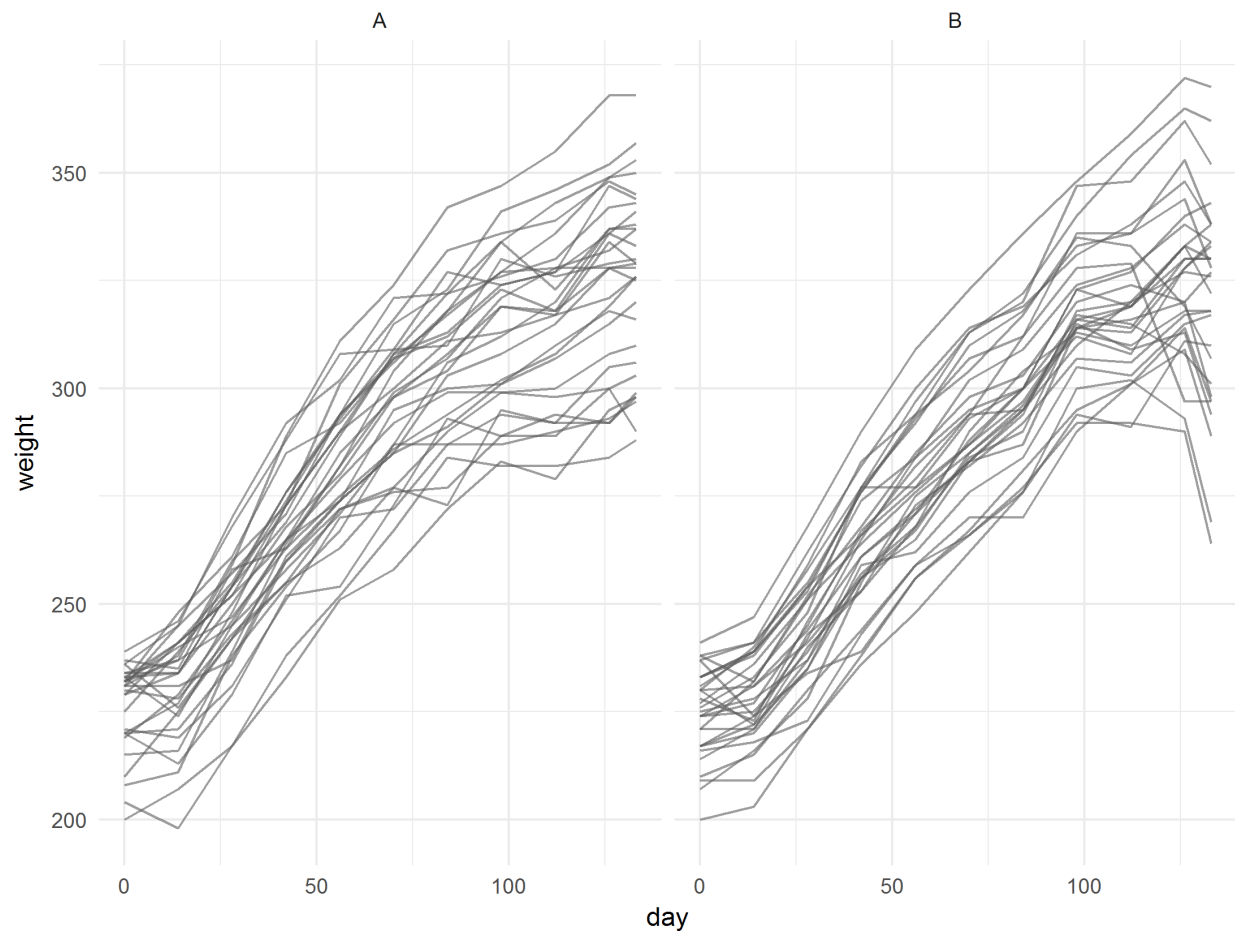
covariance matrices corresponding to treatment group using the classical likelihood ratio test, making it reasonable to study each treatment group's covariance matrix separately. Following ?, ?, ?, and ?, we analyze the data from the $N = 30$ cattle assigned to treatment group A, which we assume share 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

2

and Daniels; see ? 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 $f$, and that $f$ belong to the second-order Sobolev space:

$$W_2^2 = \left\{ f : \ f, \ f' \text{ absolutely continuous, } \int \left(f''\left(x\right)\right)^2 \ dx < \infty \right\}.$$

Figure ?? displays the he observed weight trajectories over time. Figure ?? shows the corresponding fitted mean curves.

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.

The analysis of ? concluded that the regressogram (Figure ??) and variogram (Figure ??) suggest that both sample generalised autoregressive parameters and the logarithms of the
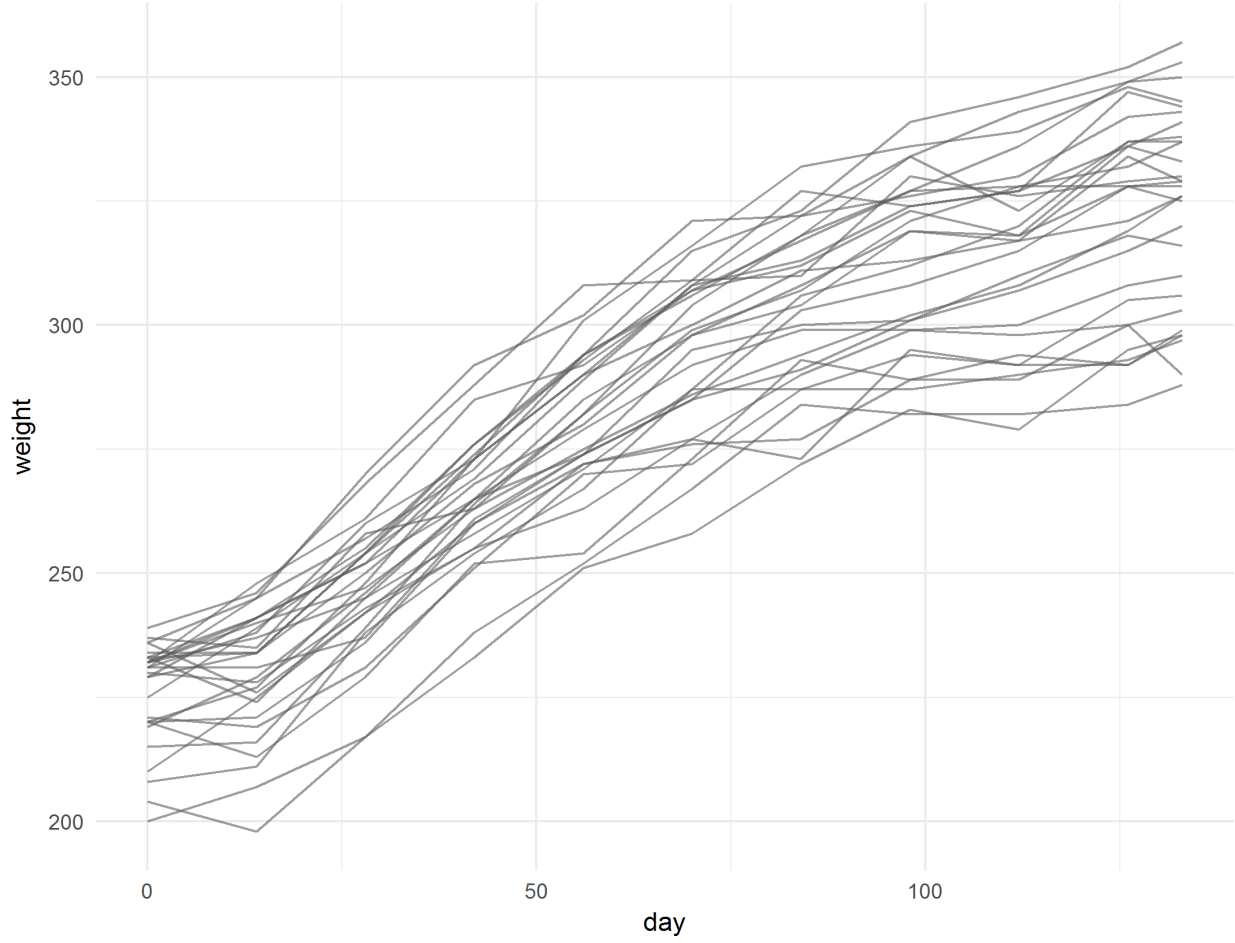
Figure 2: Weight trajectories over the observation period for experimental units in treatment group A.

innovation variances can be characterized in terms of cubic functions of the lag only. They model
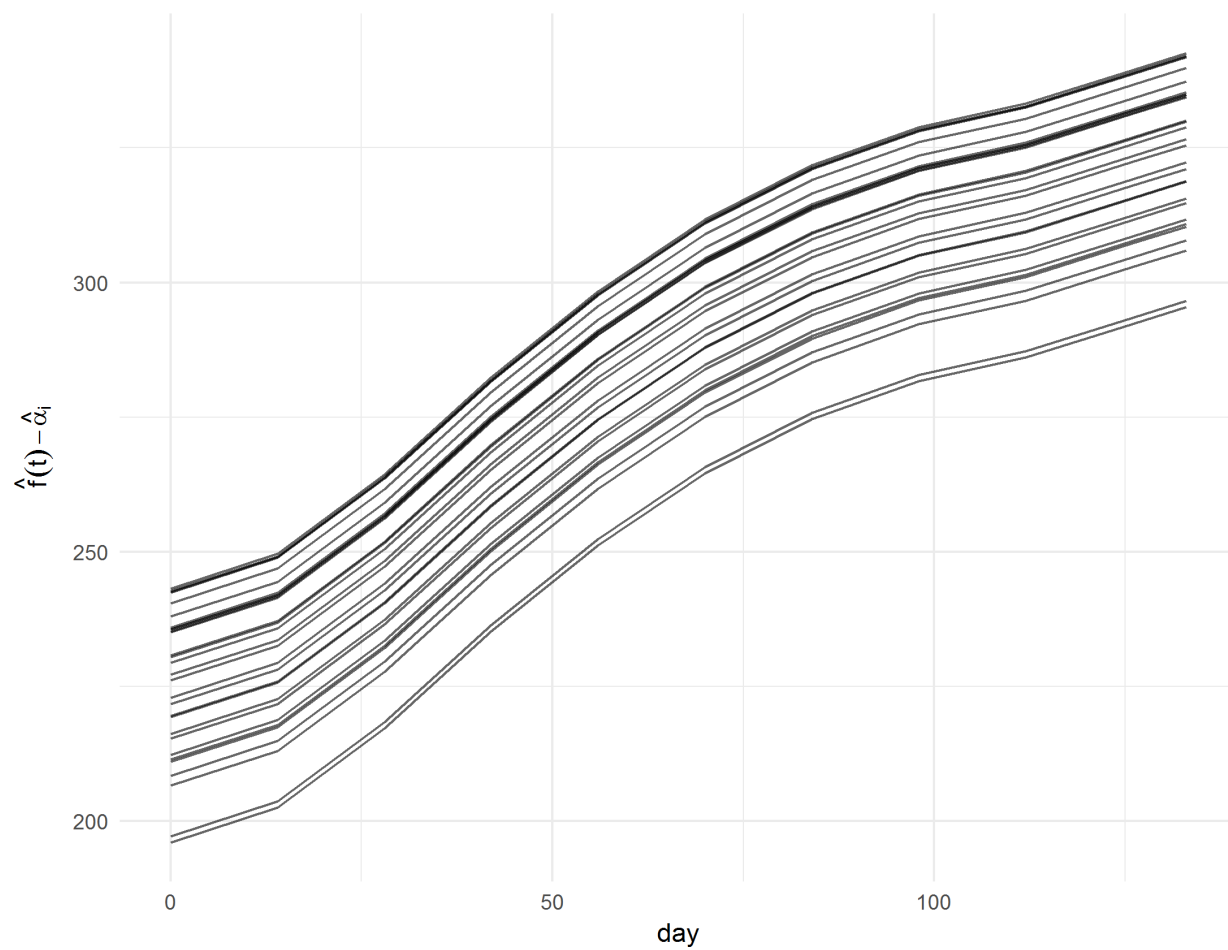
$$\phi_{ts} = x'_{ts}\gamma,$$

and

$$\log\left(\sigma_t^2\right) = z'_t\xi,$$

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

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

They estimate of $\gamma$ and $\xi$ via maximum likelihood. The cubic behaviour in the regressogram is not as obvious, however, if we examine the regressogram along each slive of $t$. Figure ??

4

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

shows the regressogram faceted along each value of $t = 1, \dots, 10.5$.

|   | s | | | | | | | | | | | $\log\left(\sigma_t^2\right)$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
|   | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 10.5 |   |
| 1 | 1 | | | | | | | | | | | 4.673 |
| 2 | 1.00 | $\ddots$ | | | | | | | | | | 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 |
| 6 | 0.04 | -0.28 | 0.11 | 0.37 | 0.82 | | | | | | | 3.363 |
| 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 | $\ddots$ | | 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 |

(The leftmost label "t" marks the row index column.)

Table 2: Cattle data: treatment group A sample generalized autoregressive parameters (below the main diagonal) and log sample innovation variances (rightmost column.)
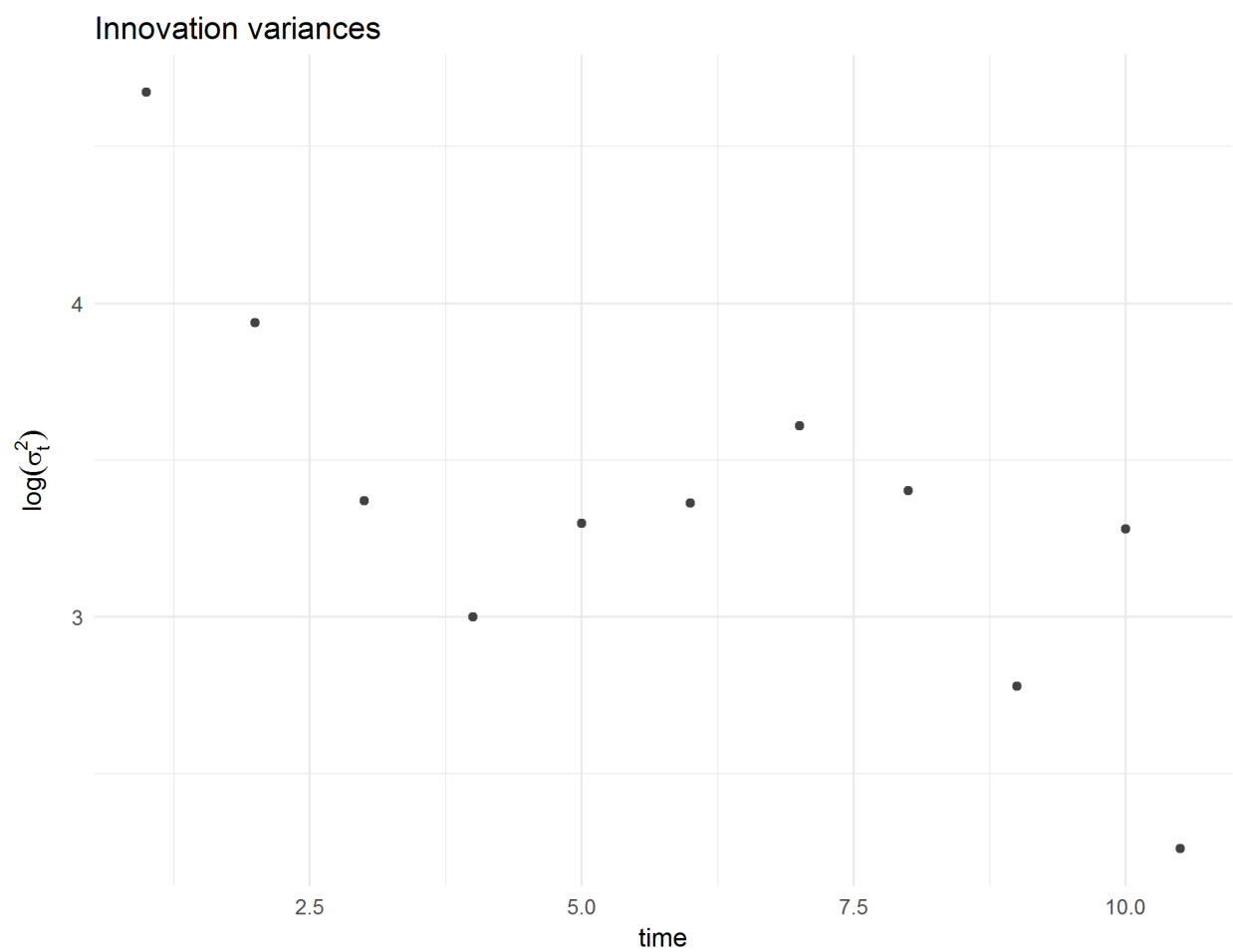
Figure 4: Sample estimates of innovation variances $\sigma_t^2$ obtained by applying the modified Cholesky decomposition to the sample covariance matrix.
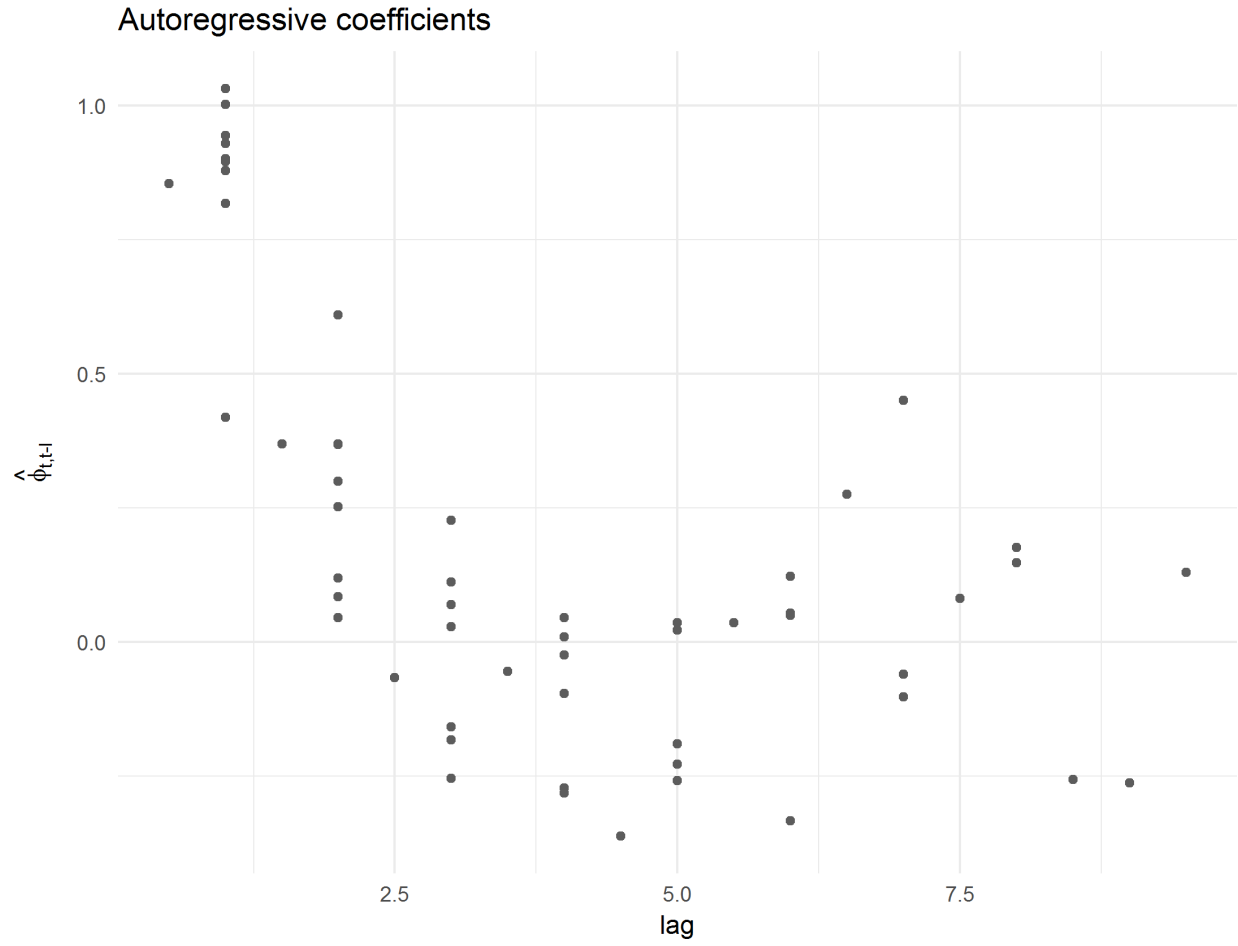
Figure 5: Sample estimates of the generalized autoregressive parameters $\phi_{ij}$ obtained by applying the modified Cholesky decomposition to the sample covariance matrix.
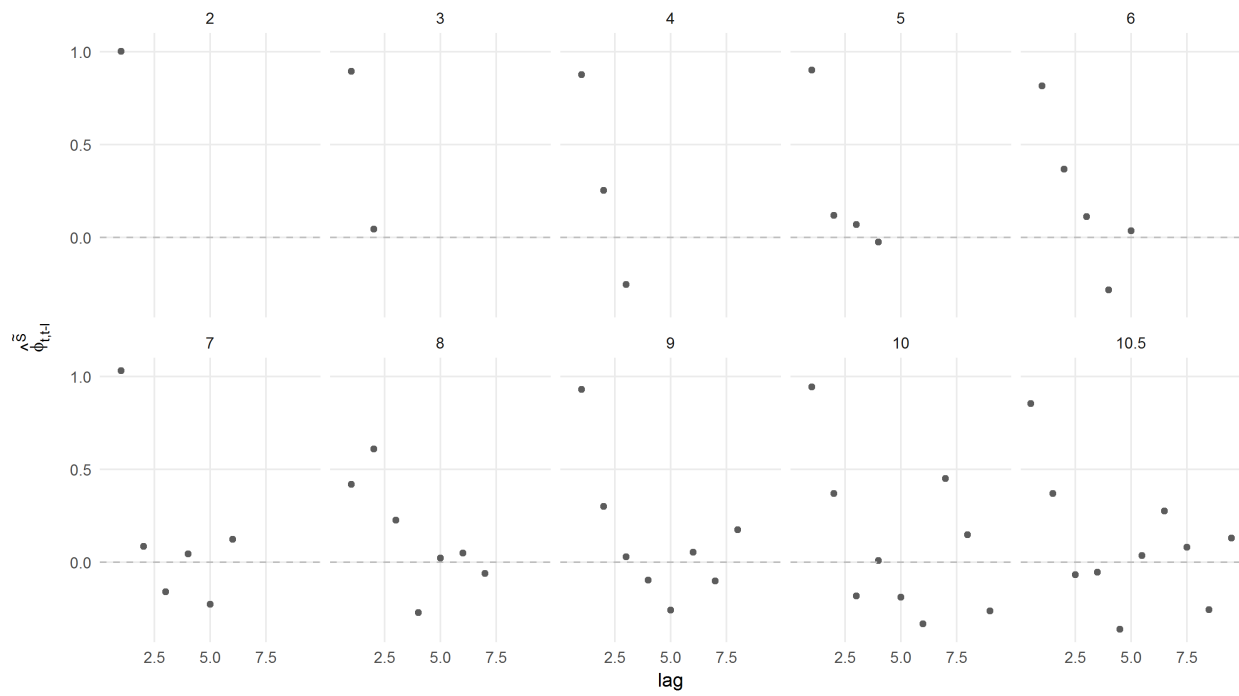
Figure 6: The cattle data regressogram sliced along the $t$ axis.