Seasonal Difference-Equation (QR-only)

Best Nth-Order with Trigonometric Seasonality

Problem. (Best Nth-Order Difference Equation with Seasonal Component and Linear Trend) You are given a scalar monthly time series $y_1, y_2, \dots, y_T$. We wish to describe and predict this sequence by an Nth-order linear difference equation augmented with a trigonometric seasonal component.

For integer orders $N \geq 0$ and $K \geq 0$, consider the model

$$y_t \approx c + d\,t + \sum_{i=1}^{N} a_i\, y_{t-i} + \sum_{k=1}^{K} \Big(\alpha_k \cos\tfrac{2\pi k t}{s} + \beta_k \sin\tfrac{2\pi k t}{s} \Big), \qquad t = N+1, \dots, T,$$

where $s = 12$ denotes the seasonal period (months) and $\beta = [\, c,\; d,\; a_1, \dots, a_N,\; \alpha_1, \dots, \alpha_K,\; \beta_1, \dots, \beta_K \,]^\top$ is a vector of unknown coefficients.

(a) Formulation as least squares. Construct the response vector $b \in \mathbb{R}^{T-N}$ and the design matrix $A \in \mathbb{R}^{(T-N)\times(2+N+2K)}$:

$$b = \begin{bmatrix} y_{N+1} \\ y_{N+2} \\ \vdots \\ y_T \end{bmatrix}, \qquad A_t = \big[1,\; t,\; y_{t-1}, \dots, y_{t-N},\; \cos\tfrac{2\pi t}{s}, \dots, \cos\tfrac{2\pi K t}{s},\; \sin\tfrac{2\pi t}{s}, \dots, \sin\tfrac{2\pi K t}{s}\big].$$

Estimate the coefficient vector β by solving the least–squares problem

$$\min_{\beta} \; \|A\beta - b\|_2,$$

using only a dense QR factorization that you implement yourself (e.g. via Householder reflectors). Compute residuals $r = A\hat{\beta} - b$ and $\mathrm{RSS} = \|r\|_2^2$.

(b) Model order selection. Repeat the fit for several values of (N, K) in user–chosen ranges, for example $N = 0, \dots, N_{\max}$ and $K = 0, \dots, K_{\max}$. For each pair (N, K), evaluate a fit score such as

$$S(N,K) = M \log\!\Big(\tfrac{\mathrm{RSS}}{M}\Big) + p\,\log M, \qquad M = T - N, \quad p = 2 + N + 2K,$$

(the Bayesian Information Criterion). Select $(N^\star, K^\star)$ minimizing $S(N,K)$.

(c) Prediction. With $(N^\star, K^\star)$ and $\hat{\beta}$ fixed, compute one–step predictions for $t = N^\star + 1, \dots, T$:

$$\widehat{y}_{t|t-1} = \hat{c} + \hat{d}\,t + \sum_{i=1}^{N^\star} \hat{a}_i\, y_{t-i} + \sum_{k=1}^{K^\star} \big(\hat{\alpha}_k \cos\tfrac{2\pi k t}{s} + \hat{\beta}_k \sin\tfrac{2\pi k t}{s} \big).$$

For forecasting h steps ahead ($h = 1, 2, \dots$), apply the same equation recursively, replacing any y_{t-i} with its forecast if $t - i > T$. The trigonometric terms are deterministic functions of t.

(d) Deliverables.

- A plot or table of $S(N,K)$ over your search grid, and the selected $(N^\star, K^\star)$.
- Estimated coefficients $\hat{c}$, $\hat{a}_1, \dots, \hat{a}_{N^\star}$, $\hat{\alpha}_1, \dots, \hat{\alpha}_{K^\star}$, $\hat{\beta}_1, \dots, \hat{\beta}_{K^\star}$.

- The in–sample mean–squared prediction error $\frac{1}{T-N^\star}\sum(y_t - \widehat{y}_{t|t-1})^2$.
- Twelve–step forecasts $\widehat{y}_{T+1}, \ldots, \widehat{y}_{T+12}$ and a plot of the data with forecasts.

Notes.

- Only dense QR decomposition may be used for the least–squares solves. No normal equations or built–in regression routines.
- The integer N controls the memory of the difference equation; K controls the number of seasonal harmonics.
- The model may be viewed as an Nth–order linear difference equation with a trigonometric forcing term. (Connections to AR or ARIMA models may be mentioned for reference, but are not required.)

Data Reference.

- The Tucson and Phoenix data are **real** and come from:

  `https://zenodo.org/records/7826348`

- The dataset **y_example.csv** is synthetic.