Discussion Paper

Models and statistical methods in `rtrim`

Jeroen Pannekoek, Patrick Bogaart, Mark van der Loo

July 2018

Summary: `rtrim` **is an r-package for the analysis of time series of counts of animal populations with missing observations. The package contains functions to estimate indices and trends and to asses the effects of covariates on these indices and trends. This report describes, in some detail, the statistical methods and models implemented in this package.**

Keywords: ***Wild life monitoring data, time series, count data, loglinear models, imputation.***

Contents

1 Introduction

The package `rtrim` (Bogaart, van der Loo and Pannekoek, 2016) is developed for the analysis of count data obtained from monitoring animal populations. Such monitoring projects typically involve a large number of sites that are surveyed annually, seasonally or monthly during some period of time. One of the principal objectives of monitoring is to assess between-year changes in abundance of the species under study. These changes are usually represented as indices, using (usually) the first year as a base year.

In practice, this kind of data often contains many missing values. This hampers the usefulness of index numbers because index numbers calculated on incomplete data will not only reflect between year changes but changes in the pattern of missing values as well. By the use of models that make assumptions about the structure of the counts, it is possible to obtain better estimates of the indices. The idea is to estimate a model using the observed counts and then to use this model to predict the missing counts. Indices can then be calculated on the basis of a completed data set with the predicted counts replacing the missing counts. The package `rtrim` implements a variety of loglinear models for this purpose.

The purpose of these models is not only to produce estimates of annual indices but also to investigate trends in these indices: is the abundance of a certain species increasing or decreasing over time. These trends need not be constant over time, allowing conclusions like “the development over time can be described by an annual increase of x% from 1980 up to 1988, no change between 1988 and 1993 and an annual decrease of y% from 1993 onwards”. TRIM also includes models that allow for effects of covariates on the trends and indices. Apart from leading to improved estimates of annual indices, covariates are also important for investigating, for instance, whether or not environmental factors such as acidification or pollution have an impact on the trends.

A problem in monitoring programmes is the oversampling of particular areas and the undersampling of others. Especially when many volunteers are involved, the more natural areas like dunes, heathland and marshes might be overrepresented whereas urban areas and farmland are underrepresented. This hinders the assessment of national figures because the changes are not necessarily similar in all area types. This situation can be remedied by the use of weights that can counter the effects of over- and undersampling.

In the application of loglinear models to the kind of data considered here, there are some statistical complications to deal with. First, the usual (maximum likelihood) approach to estimation and testing procedures for count data are based on the assumption of independent Poisson distributions (or a multinomial distribution) for the counts. Such an assumption is likely to be violated for counts of animals because the variance is often larger than expected for a Poisson distribution (overdispersion), especially when they occur in colonies. Furthermore, the counts are often not independently distributed because the counts in a particular year will also depend on the counts in the year before (serial correlation). Therefore, `rtrim` uses statistical procedures for estimation and testing that take these two phenomena into account. Second, the usual algorithms for estimating loglinear models are not practical for the large number of parameters in our models (since there is a parameter for each site the total number of parameters is larger than the number of sites which can be several hundreds). This complication

is dealt with by an algorithm that is tailor made for the applications discussed here and is much faster and requires much less memory than the usual approach.

The remaining of this report consists of the following two main sections:

Models and statistics This section gives an overview of the models and methods implemented in `rtrim` to analyse trends and estimate indices. These models belong to the class of loglinear models and, although this section is self-contained, some background in loglinear analysis will be helpful in understanding the models described here. General introductions to the theory and practice of analysing count data by loglinear models can be found in standard text books such as Agresti (1990, chapter 5), McCullagh and Nelder (1989, chapter 6), chapter 6 or Fienberg (1977). Application of loglinear models to the analysis of monitoring data, also referred to as "Poisson regression", has been discussed by ter Braak et al. (1994), Thomas (1996) and Weinreich and Oude Voshaar (1992). This section also summarizes the test-statistics implemented in `rtrim`, including goodness-of-fit tests for the models and Wald-tests for the significance of specific parameters.

Details of estimation and computation This section provides a more technical description of the estimation methods and the algorithms involved as well as more details of the calculation of the summary statistics and parameter transformations that can (optionally) be produced by `rtrim`.

2 Models and statistics

2.1 Terminology

Observed counts and missing counts The data for which the package `rtrim` is developed are counts obtained from a number of sites at a number of years (or any equi-distant time points), and optionally months (or any other season) within these years or time point. In what follows, we will, without loss of generality, speak of 'years' and 'months'. In case of annual data, the count or frequency in site i at year j will be denoted by f_{ij} $(i = 1 \ldots I, j = 1 \ldots J)$ with I the total number of sites and J the total number of years. There will usually not be observations f_{ij} for every combination of site and year and the unobserved counts are called missing counts. In case of monthly (or any other type of higher frequency) data, we will have observed frequencies f_{ijm} $(m = 1 \ldots M)$ with M the number of months.

To cover cases with and without monthly observations, we will use the notation $ij(m)$ to denote both ij, in case of yearly observations, and ijm in case of monthly observations.

Expected and estimated counts The counts are viewed as random variables. The expected counts are the expected values of the counts. The models, to be discussed in the next subsection, express the expected counts as a function of site-effects and time-effects (or, site-parameters and time-parameters). In many cases it will be possible to estimate the model parameters and hence to calculate an estimated (or predicted) expected count for every combination of i and j (and optionally m) even with a substantial number of missing counts. This depends however on the model type and the pattern of missing values. In general, complicated models with many parameters can only be estimated if the data are not too sparse (the number of missing data is not too large), and simple, but perhaps not very realistic, models can be estimated even with very sparse data. `rtrim` will inform you if a chosen model cannot be estimated because the data are too sparse. In the following, expected counts will be denoted by $\mu_{ij(m)}$, and estimated expected counts (also be called estimated counts) will be denoted by $\hat{\mu}_{ij(m)}$.

Imputed counts The count after imputation (*imputed count*) for a Site by Time combination, denoted by $f^{+}_{ij(m)}$, equals the observed count if an observation is made and equals the estimated count $\hat{\mu}_{ij(m)}$ if an observation is missing, i.e.,

$$f^{+}_{ij(m)} = \begin{cases} f_{ij(m)} & \text{if } f_{ij(m)} \text{ availabe (observed),} \\ \hat{\mu}_{ij(m)} & \text{otherwise.} \end{cases}$$

Observed, model based and imputed time-total For year j, the *observed total* is $f_{+j} = \sum_{i \in \text{obs}} f_{ij}$, where the notation $i \in$ obs denotes that summation is over available (observed) f_{ij} only. Similarly, the *model-based* total is defined as $\mu_{+j} = \sum_i \mu_{ij}$, and the *imputed total* as $f^{+}_{+j} = \sum_i f^{+}_{ij}$.

Similarly, for monthly observations, we define $f_{+j+} = \sum_{i,m \in \text{obs}} f_{ijm}$, $\mu_{+j+} = \sum_{im} \mu_{ijm}$, and $f^{+}_{+j+} = \sum_{i,m \in \text{obs}} f^{+}_{ijm}$

Model based and imputed index values An annual index value, index for short, describes the increase or decrease of a species aggregates over all sites, relative to some specific reference year or time period. Index values are computed as the total for a specific year, divided by the total for a specified reference year or period. Usually, for these indices, the first year of a time series is used as the reference year, but it is possible to select any other year to serve as reference. For the exposition in most of the remainder of this report it is assumed, however, that the first time-point is the base time-point. The model based indices are indices calculated from the model based totals and the imputed indices are indices calculated from the imputed totals.

2.2 Models

This section gives a brief description of the models that are used in `rtrim` to analyse trends and estimate indices. These models belong to the class of loglinear models. Loglinear models are linear models for the logarithm of expected counts in contingency tables (in our case the two-way Site by Time table).

Because `rtrim` models for monthly data sets are expressed as a mixture of models for yearly data, first the 'yearly' models will be presented, and then the corresponding 'monthly' models.

2.2.1 Model 1: no time-effects

A very simple, base-line, model for $\ln \mu_{ij}$ is:

$$\ln \mu_{ij} = \alpha_i, \tag{1}$$

with α_i the effect for site i. For the expected counts under this model we have $\mu_{ij(m)} = \exp(\alpha_i)$. This "no time-effects" model implies that the counts vary only across sites and not across time-points; the model based time-totals are thus equal for each time point and the model based indices are all equal to one.

2.2.2 Model 2: Linear (switching) trend

A model with a site-effect and a linear (on the log-scale) effect of time can be written as

$$\ln \mu_{ij} = \alpha_i + \beta(j-1) \tag{2a}$$

According to this model the $\ln \mu_{ij}$'s for each site i are a linear function of j with slope β; the log expected count increases with an amount β from one time-point to the next. Model (2a) can be rewritten in multiplicative form as:

$$\mu_{ij} = a_i b^{(j-1)} = b\mu_{i,j-1} \tag{2b}$$

with $a_i = \exp(\alpha_i) = \mu_{i,1}$ and $b = \exp(\beta)$ This formulation shows that for each site the expected count at some time-point j ($j > 1$) is a factor b times the expected count at the previous time-point. For the model based time-totals we have $\mu_{+j} = b^{(j-1)} \sum_i a_i$, and the model based indices are $b^{(j-1)}$.

Model (2a) implies exponential growth or decrease in the expected counts from each time point to the next. Such a model may give an adequate description of short time series but will usually become unrealistic if the time series get longer. A switching trend model allows the slope parameter to change at some time points.

For instance, a model with a slope β_1 for time points 1 to 4, a slope β_2 for time points 5 to 7 and a slope β_3 for time points beyond 7 is a switching trend model with two changes in slope, one at time point 4 and one at time point 7. The time points (4 and 7 in this example) where the slope parameter changes are called changepoints or knots and will be denoted by k_l, with $l = 1 \ldots L$ and L the number of changepoints ($k_1 = 4$, $k_2 = 7$ and $L = 2$ in this example).

This model can be reformulated to encompass the no time-effects model (1) by setting the slope to zero from the first time point up to the first changepoint, to β_1 from the first to the second changepoint and so on. The no time-effects model is then obtained if there are no changepoints and the model in the example above is obtained if we set three changepoints: $k_1 = 1$, $k_2 = 4$ and $k_3 = 7$. The linear trend model (2a) is obtained if there is a changepoint at the first time-point only.

In this formulation, the log expected counts for a model with L changepoints can be written as

$$\ln \mu_{i,j} = \begin{cases} \alpha_i & \text{for } 1 \le j \le k_1 \\ \alpha_i + \beta_1(j - k_1) & \text{for } k_1 \le j \le k_2 \\ \quad \vdots & \\ \alpha_i + \beta_1(k_2 - k_1) + \beta_2(k_3 - k_2) + \cdots + \beta_l(j - k_l) & \text{for } k_l \le j \le k_{l+1} \\ \alpha_i + \beta_1(k_2 - k_1) + \beta_2(k_3 - k_2) + \cdots + \beta_L(j - k_L) & \text{for } k_L < j \le J \end{cases}$$

So the log expected counts are constant (equal to α_i) for time points up to and including k_1. At time point $(k_1 + 1)$ the log expected count is $\alpha_i + \beta_1$. The increase between successive time points (slope) remains β_1 until the next change point k_2 is reached where the increase becomes β_2, and so on.

The equations for the log expected counts can be comprised into a single equation as follows:

$$\ln \mu_{ij} = \alpha_i + \sum_{l=1}^{L} (\beta_l - \beta_{l-1})(j - k_l)\kappa(j, k_l), \tag{3}$$

where $\beta_0 = 0$ and the function $\kappa(j, k)$ is defined by

$$\kappa(j, k) = \begin{cases} 0 & \text{for } j \le k_l \\ 1 & \text{for } j > k_l \end{cases}$$

2.2.3 Model 3: Effects for each time-point

An alternative to describing the development in time with a (number of) linear trend(s) is to use a model with separate parameters for each year. A model with effects for each site and each year can be expressed as

$$\ln \mu_{ij} = \alpha_i + \gamma_j \tag{4}$$

with γ_j the effect for time j on the log-expected counts. One restriction is needed to make the parameters of this model identifiable. In `rtrim`, the parameter γ_1 is set to zero. Model (4) can be rewritten in multiplicative form as:

$$\mu_{ij} = a_i c_j \tag{5}$$

with $a_i = \exp(\alpha_i) = \mu_{i,1}$, $c_1 = \exp(0) = 1$ and $c_j = \exp(\gamma_j)$. From (5) we have for the expected total for time j: $\mu_{+j} = \sum_i \mu_{ij} = c_j \sum_i a_i$ and so the model based indices are identical to the parameters c_j (since $\mu_{+j}/\mu_{+1} = c_j$).

The time parameters in model (4) can be decomposed in a linear trend parameter (β_j^*) and parameters (γ_j^*) describing the deviations from this linear trend for each year. Such a representation makes it easy to investigate for which years significant deviations from the linear trend occur (γ_j^* different from zero). One way of obtaining such a decomposition is by fitting a linear regression line through the $\ln \mu_{ij}$ of model (4), see section(4.2) for the details. This reparameterization can be written as

$$\ln \mu_{ij} = \alpha_i^* + \beta^* d_j + \gamma_j^* \tag{6}$$

with d_j equal to j minus the average of the j's, so $d_j = j - \frac{1}{j}\sum_j j$. The parameter α_i^* is the intercept and the parameter β^* is the slope of the regression line through the $\ln \mu_{ij}$. The parameters γ_j^* are the deviations of the $\ln \mu_{ij}$ from this regression line. Note that (6) is just a different version of (4) and (5), the expected counts and model based indices being the same for all three representations.

The model with time-point parameters is equivalent to a switching trend model when all time-points (except the last) are changepoints. For the model with time-point parameters the trend between time-points j and $j+1$ is

$$\ln \mu_{ij+1} - \ln \mu_{ij} = \gamma_{j+1} - \gamma_j \tag{7}$$

and for the equivalent switching trend model the trend is (compare (3))

$$\ln \mu_{ij+1} - \ln \mu_{ij} = \beta_j \tag{8}$$

and $\beta_1 = \gamma_2$, since $\gamma_1 = 0$.

So, the switching trend model (3) is a more general model than the time-effects model (4) since it includes this last model as a special case.

2.2.4 Extended model formulations for monthly data

In case monthly data are used, i.e. f_{ijm} instead of f_{ij}, models 1, 2, and 3 are extended to include month effects, denoted by additional parameters δ_m (for month m). In all cases, month effects are expressed similar to how year effects are expressed in Model 3, and the extended model definitions now read

$$\ln \mu_{ijm} = \alpha_i + \delta_m \tag{9}$$

$$\ln \mu_{ijm} = \alpha_i + \beta(j-1) + \delta_m \tag{10}$$

$$\ln \mu_{ijm} = \alpha_i + \gamma_j + \delta_m \tag{11}$$

or similar, for models 1, 2 and 3, respectively (note that the example given here applies to the simplest version of model 2, .i.e. Eqn (2a), but can be applied to the more generic version Eqn (3) as well). As with γ parameters, $\delta_1 \equiv 0$.

2.3 Effects of categorical covariates on the trend

Both model 2 and model 3 are restrictive in the sense that the time related parameters (β, γ and δ) are assumed to be the same for each site. By the use of covariates, this assumption can be relaxed and the models can be improved. The `rtrim` package accomodates additive effects of categorical covariates on trends and time-point parameters. For this purpose, dummy-variables are created for the categories of each covariate. Since one of the dummies is redundant, the dummy variable for the first category of each covariate is omitted. The values of these dummy

variables are denoted by $z_{ij(m)k}$, $(k = 1 \ldots K)$ with K the sum of the numbers of categories of the covariates minus the number of covariates.

An extension of the simple linear trend model (2a) that allows for additive effects of K covariates on the slope parameter is

$$\ln \mu_{ij} = \alpha_i + (\beta_0 + \sum_{k=1}^{K} z_{ijk}\beta_k)(j-1) \tag{12}$$

so that the slope of the linear trend for site i and year j consists of a for all i and j common component β_0 (which is the slope parameter for site by time combinations belonging to the first categories of all covariates) plus a component that is the sum of the effects of the categories to which site i belongs at time j. Note that the values of covariates can vary not only across sites but also across time points. This allows for the possibility that, for instance, a site is classified as 'wood' at some point in time but as 'farmland' at another point in time. A switching trend model with effects of covariates on each of the slope parameters is obtained similarly by replacing β_l in (3) with $\beta_{l0} + \sum_{k=1}^{K} z_{ijk}\beta_{lk}$.

An extension of model 3 that allows for additive effects of categorical covariates on the time-effects is:

$$\ln \mu_{ij} = \alpha_i + \gamma_{j0} + \sum_{k=1}^{K} z_{ijk}\gamma_{jk} \tag{13}$$

The effect of time j at site i now consists of a for all sites common component γ_{j0} (which is the time-effect for time j for sites belonging to the first categories of all covariates) plus an effect $\sum_k z_{ijk}\gamma_{jk}$, that is specific for the combination of categories of the covariates.

Above formulations, which are given here for yearly observations only, can be extended for monthly observations by inclusion of categorial month effects, similar to the year effects (13)

$$\ln \mu_{ijm} = \ldots + \delta_{m0} + \sum_{k=1}^{K} z_{ijmk}\delta_{mk} \tag{14}$$

2.4 Changepoints and model estimability

In many cases, users want make as few assumptions as possible regarding actual trend changes, and therefore would like to use model 3 or, equivalently, a 'maximal' model 2, where each time point is treated as a change point. However, not in all cases will there be sufficient observations to estimate the corresponding model parameters. A single year without any observations is one simple example.

In applications it will often be the case that a switching trend or time-parameters model with covariates cannot be estimated owing to a lack of observations. For the time-parameters model to be estimable, it is necessary that for each time-point there are observations for each category of each covariate. For the switching trend model to be estimable it is necessary that for each time-interval between two adjacent changepoints (time-points j for which $k_l < j \leq k_{l+1}$) there is at least one observation for each category of each covariate. `rtrim` checks these conditions and, if necessary, an error message will be issued indicating for which time-interval (time-point) and covariate category there are no observations.

An other option is, for the switching trend model, to automatically delete changepoints such that for the remaining time-intervals there are observations for each category of each covariate. This is accomplished by deleting the changepoint corresponding to the end point of the first time-interval for which no observations are available and then checking again, beginning with the newly created interval.

Note that this procedure is aimed at the identification of a 'maximum' set of change points, given the amount of actual observations. The alternative is to identify a 'minimal', parsimonious, model (a model with as few parameters as possible, without compromising the explanatory power of the model). This can be carried out by a stepwise selection of changepoints, explained in Section 4.1.

2.5 Overall trend

When covariates are used, trends and indices vary between sites and the models do not provide a measure of the trend in the aggregated (over sites) time-counts. Although the between-sites differences in trends will usually be of scientific interest since they reflect the effects of covariates on the trend, the trend in the aggregated time-counts will often also be of interest since this 'overall trend' reflects changes in the total population over time. A simple measure of overall trend can be obtained as the ordinary least squares (ols) estimator of the slope parameter, β_+ say, of a linear regression line through the log estimated model-based time-totals, $\ln \hat{\mu}_{+j}$. Thus, as the ols estimator $\hat{\beta}_+$ of β_+ in the expression

$$\ln \hat{\mu}_{+j} = \alpha + \beta_+(j-1) + \varepsilon_j \tag{15}$$

with ε_j the deviation of the log estimated time-total for time j from the linear trend.

To obtain expressions for the ols-estimators of the slope parameters, we introduce the following notation: $\boldsymbol{X}_1$ is a J-vector (a vector of length J) with all elements equal to 1, $\boldsymbol{X}_2$ a J-vector with values $j-1$ $(j=1,\ldots,J)$, $\boldsymbol{X}=(\boldsymbol{X}_1,\boldsymbol{X}_2)$ and $\boldsymbol{y}$ a J-vector with values $\ln \mu_{+j}$. Then we have for the ols-estimators for the intercept and slope in model (2.5):

$$\hat{\boldsymbol{\beta}} = (\alpha,\beta_+)^T = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\mathbf{y}, \tag{16}$$

It is important to note that the estimator $\hat{\beta}_+$ of the overall slope is not viewed as an estimator of a parameter of a model thought to have generated the $\ln \hat{\mu}_{+j}$'s but as a descriptive statistic highlighting one aspect (the linear trend) of the $\ln \hat{\mu}_{+j}$'s. The $\ln \hat{\mu}_{+j}$'s in (15) are estimates that can have been derived from any of the models discussed before, and will not generally follow a linear trend.

Although $\hat{\beta}_+$ is defined by ols-regression, its variance is estimated in a way that is different from the usual ols-regression approach. In line with the interpretation of $\hat{\beta}_+$ a summary statistic (function) of the $\ln \hat{\mu}_{+j}$'s, estimator of its variance is obtained from the estimated covariance matrix of the $\ln \hat{\mu}_{+j}$'s, which in turn is derived from the estimated covariance matrix of the parameters of the model used to generate the $\ln \hat{\mu}_{+j}$'s (see, section 5).

2.6 Using weights

In some instances it is advisable to use cell weights to improve the estimates of national indices, see van Strien et al. (1995) for an example. For instance, if sites from urban areas are underrepresented relative to sites from other areas, weights could be calculated such that the weighted total surface of urban sites equals the population total surface of urban areas and the weighted total surface of other areas also equals the corresponding population surface. Then, assuming that the counts are proportional to the surface of the sites, the counts can be multiplied by these weights to obtain a better representation of the population counts. More generally, weights can be determined such that the weighted total surface of sites of a certain type at a certain point in time equals, or is proportional to, the total population surface of sites of that type. This kind of weighting can counter the effects of over- and undersampling and is easy to incorporate in the loglinear modelling approach.

When weights are used, interest will be in models describing the weighted expected counts. If the weights are denoted by $w_{ij(m)}$, the expected value of the weighted counts will be $\mathrm{E}[w_{ij(m)}f_{ij(m)}] = w_{ij(m)}\mu_{ij(m)}$ since the weights are known constants. A model, for instance model 3 (effects for each time-point), for the weighted expected counts can be written as

$$\ln w_{ij}\mu_{ij} = \alpha_i + \gamma_j, \tag{17}$$

or

$$w_{ij}\mu_{ij} = a_i c_j. \tag{18}$$

This model implies for the unweighted expected counts

$$\ln \mu_{ij} = \alpha_i + \gamma_j - \ln w_{ij}. \tag{19}$$

The $\ln w_{ij}$ are parameters that are known in advance. Such parameters are called an offset in the terminology of generalized linear models (glm's) (McCullagh and Nelder, 1989).

When weights are used, the model based indices are $\sum_i w_{ij}\mu_{ij} / \sum_i w_{i1}\mu_{i1}$ (assuming the first time point is taken as reference). These indices will not change if the weights are multiplied by a constant different from zero, but the model based totals for the time-points will change. If the weights do not change over time we can write $w_{ij} = w_i$, with w_i the common weight for all time-points for site i. The indices for model (18) can then be expressed as $\sum_i w_i a_i c_j / \sum_i w_i a_i = c_j$ showing that the indices are independent of the weights and the weighted model based indices are equal to the unweighted model based indices. More generally, weighted and unweighted model based indices are equal if the weights are equal for all time-points and the time related parameters are the same for all sites. Thus, if $w_{ij} = w_i$, the weighting does not affect the indices for models without covariates but does affect the indices if covariates are used.

Weighted model based indices will be calculated using the weighted estimated counts and weighted imputed indices will be calculated using the weighted observed counts $w_{ij}f_{ij}$ if they are available and the weighted estimated counts otherwise.

The weights as described in this subsection are part of the model, they are multiplicative factors used to increase/decrease counts for site/time combinations that are underrepresented/overrepresented in the sample and do not change the variances of the observations. This specific type of weighting should not be confused with the weighting as performed by estimation methods such as weighted least squares or generalisations thereof such

as the iterative weighted least squares algorithm used for generalised linear models. In such procedures the observations are weighted by the inverse of their variances and the weights are part of the estimation procedure but not of the model.

In case of monthly observations, this reasoning does not change, and e.g. Equations (17) and (19) are written as

$$\ln w_{ijm}\mu_{ijm} = \alpha_i + \gamma_j + \delta_m \tag{20}$$

and

$$\ln \mu_{ijm} = \alpha_i + \gamma_j + \delta_m - \ln w_{ijm}. \tag{21}$$

2.7 Estimation options

The usual approach to statistical inference for loglinear models is to use maximum likelihood (ML) estimation and associated calculations of standard errors and test statistics. These estimation and testing procedures are based on the assumption of independent Poisson distributions (or a multinomial distribution) for the counts. Such an assumption is likely to be violated for counts of animals because the variance is often larger than expected for a Poisson distribution (overdispersion), especially when they occur in colonies. Furthermore, the counts are often not independently distributed because the counts at a particular point in time will often depend on the counts at the previous time-point (serial correlation). The `rtrim` package uses procedures for estimation and testing that take these two phenomena into account (a Generalised Estimating Equations (GEE) approach, see section 3.2 for details). This procedure is based on the following assumptions for the variance of the counts and the correlation between the counts for adjacent time-points:

$$\text{var}(f_{ij}) = \sigma^2\mu_{ij} \tag{22}$$

and

$$\text{cor}(f_{ij}, f_{i,j+1}) = \rho \tag{23}$$

The parameter σ^2 is called a dispersion parameter. For $\sigma^2 = 1$, the variance of f_{ij} is equal to its expectation which is the variance under the Poisson assumption. The parameter ρ is the serial correlation parameter. The counts are independent if $\rho = 0$. If both $\sigma^2 = 1$ and $\rho = 0$, the estimation procedure used in `rtrim` is identical to the usual maximum likelihood approach. If $\sigma^2 \neq 1$ and $\rho = 0$, the estimates of parameters (and expected counts and indices) are equal to the maximum likelihood estimates but the estimated standard errors and test statistics will be different. If $\rho \neq 0$ both the estimates of parameters and standard errors differ from the maximum likelihood estimates. The difference between GEE and ML estimates of parameters is usually small and tends to decrease as the counts increase. However, the corresponding difference between estimated standard errors and test-statistics need not be small nor decreases when the counts become larger. So, allowing ρ and σ^2 to be unequal to 0 and 1 respectively, has little impact on the estimated parameters but can have important effects on standard errors. In `rtrim` options can be set that allow the user to specify whether overdispersion and/or serial correlation must be taken into account or not. If either of these options is used estimates of σ^2 and/or ρ will be calculated and used in estimation and testing procedures.

In case of monthly observations, overdispersion is allowed, and Equation (22) is written as

$$\text{var}(f_{ijm}) = \sigma^2\mu_{ijm}$$

but serial correlation is not considered ($\rho \equiv 0$) due to the complexities associated with intra-annual serial correlations in species abundance.

2.8 Test-statistics

2.8.1 Model goodness-of-fit tests

The goodness-of-fit of loglinear models is generally tested by *Pearson's chi-squared statistic*, given by

$$\chi^2 = \sum_{ij(m)} \frac{(f_{ij(m)} - \hat{\mu}_{ij(m)})^2}{\hat{\mu}_{ij(m)}} \tag{24}$$

or by the *likelihood ratio test* given by

$$LR = 2 \sum_{ij(m)} f_{ij(m)} \ln\left(\frac{f_{ij(m)}}{\mu_{ij(m)}}\right) \tag{25}$$

where the summation is over observed (i, j) or (i, j, m) only. For independent Poisson observations, both statistics are asymptotically χ^2_ν distributed, with ν the number of degrees of freedom (equal to the number of observed counts minus the number of estimated parameters). Models are rejected for large values of these statistics and small values of the associated significance probabilities. These tests indicate how well the model describes the observed counts.

The likelihood ratio statistic can be used to test for the difference between nested models. That is, if we have two models, M_1 with p parameters and M_2 with the same p parameters plus q additional parameters, then M_1 is said to be nested within M_2 (M_1 can be obtained from M_2 by setting the q additional parameters of M_2 equal to zero). Now, model M_1 can be tested against model M_2 by using the difference between the likelihood ratio statistics for the two models ($LR_{1-2} = LR_1 - LR_2$, say) as test statistic. This difference is also a likelihood ratio statistic and therefore asymptotically χ^2_ν distributed, with degrees of freedom ν equal to the difference in degrees of freedom for the two models which is also equal to the number of additional parameters q.

Another approach to comparing models is by the use of Akaike's Information Criterion (AIC) (see, e.g. McCullagh and Nelder (1989), page 91). For loglinear models this criterion can be expressed as $C + LR - 2\nu$ where the constant C is the same for all models for the same data set. According to this approach, models with smaller values of AIC, or equivalently $LR - 2\nu$, provide better fits than models with larger values. Contrary to comparing models by using the likelihood ratio test for the difference, comparing models on the basis of AIC-values is not restricted to nested models.

If the counts are not (assumed to be) independent Poisson observations and either σ^2 or ρ is estimated, the statistics defined by (24) and (25) are not asymptotically χ^2_ν distributed and the associated significance probabilities are incorrect. Also, the AIC cannot be used for comparing models. However, Wald-tests (to be described below) can still be used to test for the significance of (groups of) parameters.

2.8.2 Wald-tests for significance of parameters

A number of tests can be performed in `rtrim` to test for the significance of groups of parameters. These so called Wald-tests are based on the estimated covariance matrix of the parameters and since this covariance matrix takes the overdispersion and serial correlation into account (if specified), these tests are valid, not only if the counts are assumed to be independent Poisson observations but also if σ^2 and/or ρ is estimated. The general form of the Wald-statistic for testing simultaneously whether several parameters are different from zero is

$$W = \hat{\boldsymbol{\theta}}^T \left[\text{var}(\hat{\boldsymbol{\theta}})\right]^{-1} \hat{\boldsymbol{\theta}},$$

with $\hat{\boldsymbol{\theta}}$ a vector containing the parameter estimates to be tested and $var(\hat{\boldsymbol{\theta}})$ the covariance matrix of $\hat{\boldsymbol{\theta}}$.

The following Wald-tests can be performed in `rtrim`:

1. Test for the significance of the slope parameter (model 2).
2. Tests for the significance of changes in slope (model 2).
3. Test for the significance of the deviations from a linear trend (model 3).
4. Tests for the significance of the effect of each covariate (models 2 and 3).

Wald-tests are asymptotically χ^2_v distributed, with the number of degrees of freedom equal to the rank of the covariance matrix var($\hat{\boldsymbol{\theta}}$). The hypothesis that the tested parameters are zero is rejected for large values of the test-statistic and small values of the associated significance probabilities (denoted by p), so parameters are significantly different from zero if p is smaller than some chosen significance level (customary choices are 0.01, 0.05 and 0.10)

In addition to these tests the significance of each individual parameter can be tested by a t-test e.g., a parameter is significantly (at the 0.05 significance level) different from zero if it exceeds plus or minus 1.96 times its standard error.

2.9 Equality of model based and imputed indices

For the model with parameters for each time point (model 3, i.e., without month effects), the model-based and imputed indices are equal if $\rho = 0$ and no weighting is used. This is explained in this subsection.

Model 3 (without covariates) is the model of independence in a two-way contingency table. It is well known (e.g. Fienberg (1977, ch. 2) that if the parameters of this model are estimated by maximum likelihood, the estimated expected counts satisfy

$$\sum_{i \in \text{obs}} \hat{\mu}_{ij} = \sum_{i \in \text{obs}} f_{ij} = f_{+j}, \tag{26}$$

where again the summation is over observed (i, j) only. Thus, the time-totals of the estimated expected counts, where the summation is over the observed cells only, are equal to the time-totals of the observed counts (also summing over the observed cells only, of course). For the imputed time-totals we then have

$$\sum_i f^+_{ij} = \sum_{i \in \text{obs}} f_{ij} + \sum_i \hat{\mu}_{ij} - \sum_{i \in \text{obs}} \hat{\mu}_{ij} = f_{+j} + \hat{\mu}_{+j} - f_{+j} = \hat{\mu}_{+j} \tag{27}$$

So, the imputed time-totals are equal to the estimated model based time-totals and the imputed and model based indices will both be equal to the estimates of the parameters c_j. This equality

between imputed and model based indices holds also when covariates are used since then equalities analogous to (26) and (27) apply to the imputed and model based time-totals for each group of sites sharing the same covariate values. Therefore, the imputed and model based time-totals for all sites, obtained by adding the per group time totals, must also be equal.

Equality between imputed and model based indices also holds if $\sigma \neq 1$ and $\rho = 0$ because the estimates of parameters (and expected counts) are then equal to the maximum likelihood estimates (see section 2.7) but the equality does not hold (in general) if either I) the model is not the time-effects model or II) weighting is used or III) $\rho \neq 0$.

3 Details of estimation and computation

3.1 Matrix formulation

To facilitate the discussion of estimators for the model parameters, it is convenient to formulate the models in matrix notation. If we collect the expected frequencies in an $IJ(M)$-vector $\boldsymbol{\mu} = (\mu_{1,1(,1)}, \ldots, \mu_{I,J(,M)})$, all models can be written as

$$\ln \boldsymbol{\mu} = \boldsymbol{A\alpha} + \boldsymbol{B\beta}, \tag{28}$$

with $\boldsymbol{\alpha}$ a vector of length I containing the site-parameters and $\boldsymbol{\beta}$ a vector containing the time related parameters (which can be either all β_j or γ_j, augmented with all δ_m in case of monthly data). $\boldsymbol{A}$ and $\boldsymbol{B}$ are dummy matrices for the site-effects and time-effects. For all models in `rtrim`, $\boldsymbol{\alpha}$ has length $p_a = I$, and $\boldsymbol{A}$ is an $IJ(M) \times I$-matrix with I dummy-variables, one for each site. The matrix $\boldsymbol{B}$ and vector $\boldsymbol{\beta}$ (of length p_b) are specific for each model.

The parameter vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ can be combined to one vector $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$ and the design matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be combined to one $IJ(M) \times p$ design matrix $\boldsymbol{X} = (\boldsymbol{A}, \boldsymbol{B})$ with $p = p_a + p_b$ the total number of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ parameters to estimate.

The model (28) can then be written as

$$\ln \boldsymbol{\mu} = \boldsymbol{X\theta} \tag{29}$$

and models for weighted counts can be specified as

$$\ln \operatorname{diag}(\boldsymbol{w})\boldsymbol{\mu} = \boldsymbol{X\theta}$$

or

$$\ln \boldsymbol{\mu} = \boldsymbol{X\theta} - \ln \boldsymbol{w}$$

with $\boldsymbol{w}$ an $IJ(M)$-vector containing the cell weights and $\operatorname{diag}(\boldsymbol{w})$ a diagonal matrix with $\boldsymbol{w}$ on the diagonal.

3.2 Generalized estimating equations

The estimation method used in `rtrim` is based on generalized estimating equations (GEE) see, Liang and Zeger (1986), Zeger and Liang (1986), McCullagh and Nelder (1989), chapter 9. Contrary to maximum-likelihood (ML) this method doesn't require the distribution of the observations to be specified in full. The specification (up to some unknown parameters) of the first two moments (expectation and covariance matrix) is sufficient. This makes it relatively easy to take overdispersion and serial correlation into account. Furthermore, the GEE approach to estimating loglinear models reduces to the usual maximum likelihood approach if the covariance matrix of the observations equals the covariance matrix of independent Poisson observations ($\sigma^2 = 1$ and $\rho = 0$).

For estimating the parameters only the observed counts can be used and therefore, in this subsection, the vector $\boldsymbol{f}$ refers to the O-vector ($O \leq IJ(M)$) with observed counts only and

similarly the rows of the matrices $\boldsymbol{X}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ corresponding with missing counts are deleted such that the dimensions of these matrices are now $O \times p$, $O \times p_a$ and $O \times p_b$, respectively.

Given the values of the parameters in the covariance matrix, the GEE estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is the solution of the estimating equation

$$\boldsymbol{U}(\hat{\boldsymbol{\theta}}) = \boldsymbol{D}^T\boldsymbol{V}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}) = \boldsymbol{0}, \tag{30}$$

with $\boldsymbol{D}$ the $O \times p$ matrix $\partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}$ and $\boldsymbol{V}$ the covariance matrix of $\boldsymbol{f}$. Since the elements of $\boldsymbol{D}$ are given by $D_{ij,k} = \partial\mu_{ij}/\partial\theta_k = X_{ijk}\mu_{ij}$ we can write $\boldsymbol{D} = \text{diag}(\boldsymbol{\mu})\boldsymbol{X}$ and for the estimating function $\boldsymbol{U}(\boldsymbol{\theta})$ we have

$$\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{X}^T \text{diag}(\boldsymbol{\mu})\boldsymbol{V}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}) \tag{31}$$

If the counts were assumed to be independently Poisson distributed, $\boldsymbol{V}$ would be $\text{diag}(\boldsymbol{\mu})$ and the function $\boldsymbol{U}(\boldsymbol{\theta})$ would reduce to $\boldsymbol{X}^T(\boldsymbol{f} - \boldsymbol{\mu})$ which is well known to be the score-function (derivative w.r.t. $\boldsymbol{\theta}$) of the likelihood associated with this assumption.

The expected value of the derivative matrix $\partial\boldsymbol{U}(\boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$ is

$$-\boldsymbol{D}^T\boldsymbol{V}^{-1}\boldsymbol{D} = -\boldsymbol{i}(\boldsymbol{\theta}), \tag{32}$$

where $\boldsymbol{i}(\boldsymbol{\theta})$ plays the same role as the Fisher information matrix for likelihood functions. In particular, if the model is correct and the observed counts are large, the distribution of the GEE estimator $\hat{\boldsymbol{\theta}}$ is approximately normal with covariance matrix $\boldsymbol{i}(\boldsymbol{\theta})^{-1}$.

For given values of the correlation and dispersion parameters, the GEE estimator for $\boldsymbol{\theta}$ (the solution to (30)) is usually obtained by Fisher scoring iterations given by

$$\begin{aligned}\boldsymbol{\theta}_{t+1} &= \boldsymbol{\theta}_t + \boldsymbol{i}(\boldsymbol{\theta}_t)\boldsymbol{U}(\boldsymbol{\theta}_t)^{-1} \\ &= \boldsymbol{\theta}_t + (\boldsymbol{D}_t^T\boldsymbol{V}_t^{-1}\boldsymbol{D}_t)^{-1}\boldsymbol{D}_t^T\boldsymbol{V}_t^{-1}(\boldsymbol{f} - \boldsymbol{\mu}_t)\end{aligned} \tag{33}$$

where t is the iteration number and $\boldsymbol{\theta}_t$, $\boldsymbol{V}_t$, $\boldsymbol{D}_t$ and $\boldsymbol{\mu}_t$ are estimates at iteration t. If $\boldsymbol{V} = \text{diag}(\boldsymbol{\mu})$ (the Poisson assumption), the current estimate of $\boldsymbol{V}$ would be $\text{diag}(\boldsymbol{\mu}(\boldsymbol{\theta}_t))$ and depend on the current estimate of $\boldsymbol{\theta}$ only. In our applications we are often not willing to assume that $\boldsymbol{V} = \text{diag}(\boldsymbol{\mu})$ because it is likely that overdispersion and serial correlation are present and $\boldsymbol{V}$ will depend on $\boldsymbol{\mu}$ as well as on dispersion and correlation parameters and estimates of these parameters are required in order to update $\boldsymbol{\theta}$. Consequently, the algorithm iterates between updating $\boldsymbol{\theta}$ and estimating the dispersion and correlation parameters as described in section 3.3.

A problem with the updating equation (33) is the size ($p \times p$) of the matrix $\boldsymbol{D}_t^T\boldsymbol{V}_t^{-1}\boldsymbol{D}_t$. The number of parameters p is at least equal to the number of sites I, which can be well over 1000. Inverting such large matrices is very time and memory consuming, and a potential source for numerical instability. The matrix $\boldsymbol{V}$ is of course even larger ($O \times O$), but for this matrix a block diagonal structure is assumed (section 3.3) which reduces the problem to inverting the covariance matrices for the observations for each site separately. As an alternative to (33) an algorithm can be applied that uses the derivatives of $\boldsymbol{U}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\beta}$ only. This procedure leads to an algorithm that is much faster and less memory consuming than an algorithm based on (33) and is described in section (3.4).

3.3 Estimation of the covariance matrix

To allow for overdispersion and serial correlation, the $O \times O$ covariance matrix $\boldsymbol{V}$ of $\boldsymbol{f}$ is expressed as

$$\boldsymbol{V} = \sigma^2 \sqrt{\text{diag}\,\boldsymbol{\mu}}\boldsymbol{R}\sqrt{\text{diag}\,\boldsymbol{\mu}}, \tag{34}$$

with $\boldsymbol{R}$ a correlation matrix. In case of no serial correlation, $\boldsymbol{R}$ simply is the identity matrix, implicating that $\boldsymbol{V} = \sigma^2\,\text{diag}\,\boldsymbol{\mu}$ in this case, or even $\boldsymbol{V} = \text{diag}\,\boldsymbol{\mu}$ when overdispersion is also absent (Poisson assumption)

A simple correlation matrix $\boldsymbol{R}$ that reflects serial correlation is obtained by assuming that within each site there is a constant correlation, ρ say, between the observed counts at years j and $j-1$ and that counts from different sites are uncorrelated. This leads to a block diagonal correlation matrix of the form

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 & & & & \\ & \ddots & & & \\ & & \boldsymbol{R}_i & & \\ & & & \ddots & \\ & & & & \boldsymbol{R}_I \end{bmatrix} \tag{35}$$

with $\boldsymbol{R}_i$ the $O_i \times O_i$ correlation matrix of the O_i observations in site i. If there are no missing values in a site i then $\boldsymbol{R}_i$ is a $J \times J$ matrix and can be expressed as

$$\boldsymbol{R}_i = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{J-1} \\ \rho & 1 & \rho & \dots & \rho^{J-2} \\ \vdots & \vdots & \vdots & & \vdots \\ \rho^{J-1} & \rho^{J-2} & \rho^{J-3} & \dots & 1 \end{bmatrix} \tag{36}$$

which reflects a declining correlation between counts as they are further apart in time. For sites with missing values the correlation matrix can be obtained from (36) by deleting the rows and columns corresponding to the time-points for which there are no observations.

Following Liang and Zeger (1986), an estimates of σ^2 can be obtained from the Pearson residuals

$$r_{ij(m)} = (f_{ij(m)} - \mu_{ij(m)})/\sqrt{\mu_{ij(m)}}, \tag{37}$$

(which are obviously only available for the O observed i, j, m combinations), as

$$\hat{\sigma}^2 = \frac{1}{O-p}\sum_{ij(m)} r_{ij(m)}^2 \tag{38}$$

where the summation is again over the observed $(i, j(, m))$ only. Note that the inclusion of p in the denominator of (38) is to account for the effect of parameter-dependency of the $\mu_{ij(m)}$'s and hence the $r_{ij(m)}$'s on the available degrees of freedom.

Similarly, an estimate of ρ can be obtained as

$$\hat{\rho} = \frac{1}{N\hat{\sigma}^2}\sum_{i=1}^{I}\sum_{j=1}^{J-1} r_{i,j}r_{i,j+1} \tag{39}$$

where the j-summation is only over consecutive pairs $j, j+1$ if both are observed, and N is the total number of all such pairs. Note again, that serial correlation is only used when observations are on an annual time scale.

3.4 An efficient algorithm

Using the partitioning $\boldsymbol{\theta} = (\boldsymbol{\alpha}^T, \boldsymbol{\beta}^T)^T$ of the parameter vector and the corresponding partitioning $\boldsymbol{X} = (\boldsymbol{A}, \boldsymbol{B})$, the estimating equation $\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{0}$ can be expressed in two equations as

$$\boldsymbol{U}_a = \boldsymbol{A}^T \operatorname{diag} \boldsymbol{\mu} \boldsymbol{V}^{-1} (\boldsymbol{f} - \boldsymbol{\mu}) = \boldsymbol{0} \tag{40a}$$

$$\boldsymbol{U}_b = \boldsymbol{B}^T \operatorname{diag} \boldsymbol{\mu} \boldsymbol{V}^{-1} (\boldsymbol{f} - \boldsymbol{\mu}) = \boldsymbol{0} \tag{40b}$$

The negative expected derivative matrix $\boldsymbol{i}(\boldsymbol{\theta})$, defined in (32), can be partitioned similarly leading to

$$\boldsymbol{i}(\boldsymbol{\theta}) = -\left[\begin{array}{c|c} \partial \boldsymbol{U}_a / \partial \boldsymbol{\alpha}^T & \partial \boldsymbol{U}_a / \partial \boldsymbol{\beta}^T \\ \hline \partial \boldsymbol{U}_b / \partial \boldsymbol{\alpha}^T & \partial \boldsymbol{U}_b / \partial \boldsymbol{\beta}^T \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{A}^T \boldsymbol{\Omega} \boldsymbol{A} & \boldsymbol{A}^T \boldsymbol{\Omega} \boldsymbol{B} \\ \hline \boldsymbol{B}^T \boldsymbol{\Omega} \boldsymbol{A} & \boldsymbol{B}^T \boldsymbol{\Omega} \boldsymbol{B} \end{array}\right] \tag{41}$$

with $\boldsymbol{\Omega} = \operatorname{diag} \boldsymbol{\mu} \boldsymbol{V}^{-1} \operatorname{diag} \boldsymbol{\mu}$.

The equations (40a) and (40b) can be solved in two steps. First we solve (40a) with respect to $\boldsymbol{\alpha}$ using the value for $\boldsymbol{\beta}$ from the previous iteration and substitute the resulting value $\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$, say in (40b), leading to

$$\boldsymbol{U}_b^* = \boldsymbol{U}_b(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \boldsymbol{\beta}). \tag{42}$$

Second, we solve (42) with respect to $\boldsymbol{\beta}$. With the new value for $\boldsymbol{\beta}$ the two steps can be repeated. This process is iterated until convergence. The resulting estimates for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ solve the equations (40a, 40b) and hence $\boldsymbol{U}(\boldsymbol{\theta}) = \boldsymbol{0}$. This two-step procedure is similar to the "concentrated likelihood" approach for solving likelihood equations (see Amemiya, 1985, Ch. 4.2.5).

To solve (40a) for $\boldsymbol{\alpha}$ we note that the matrix $\boldsymbol{A}$ contains dummy variables for each site and the matrix $\boldsymbol{V}$ is a block diagonal covariance matrix of the same form as (35) so that for site i we can write

$$\boldsymbol{1}_{O_i}^T \operatorname{diag}(\boldsymbol{\mu}_i) \boldsymbol{V}_i^{-1} (\boldsymbol{f}_i - \boldsymbol{\mu}_i) = \boldsymbol{0} \tag{43}$$

with O_i the number of observed counts for site i, $\boldsymbol{1}_{O_i}$ an O_i-vector with ones and $\boldsymbol{f}_i$ the O_i-vector with observed counts for site i with expectation $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{V}_i$. For $\boldsymbol{\mu}_i$ we can write $\boldsymbol{\mu}_i = a_i \exp(\boldsymbol{B}_i \boldsymbol{\beta})$, with $\boldsymbol{B}_i$ the matrix with the rows of $\boldsymbol{B}$ corresponding to the observations in site i. Now, (43) can be written as

$$\boldsymbol{\mu}_i^T \boldsymbol{V}_i^{-1} (\boldsymbol{f}_i - \hat{a}_i \exp(\boldsymbol{B}_i \boldsymbol{\beta})) = \boldsymbol{0}$$

leading to

$$\hat{a}_i = \boldsymbol{\mu}_i^T \boldsymbol{V}_i^{-1} \boldsymbol{f}_i / \boldsymbol{\mu}_i^T \boldsymbol{V}_i^{-1} \exp(\boldsymbol{B}_i \boldsymbol{\beta}). \tag{44}$$

To solve the equation $\boldsymbol{U}_b^* = \boldsymbol{0}$ for $\boldsymbol{\beta}$ a Fisher scoring algorithm analogous to (33) can be used. The expected value of the required derivative matrix, $\boldsymbol{i}_b^*$ say, can be written as

$$-\boldsymbol{i}_b^* = \partial \boldsymbol{U}_b^* / \partial \boldsymbol{\beta}^T = \partial \boldsymbol{U}_b / \partial \boldsymbol{\beta}^T + \left(\partial \boldsymbol{\alpha}^T / \partial \boldsymbol{\beta}\right)\left(\partial \boldsymbol{U}_b^T / \partial \boldsymbol{\alpha}\right) \tag{45}$$

where the derivatives are evaluated in $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$.

Next, differentiating both sides of the equation $\boldsymbol{U}_a(\hat{\boldsymbol{\alpha}}(\boldsymbol{\beta}), \boldsymbol{\beta}) = \boldsymbol{0}$ with respect to $\boldsymbol{\beta}$ we obtain

$$\left(\partial \boldsymbol{\alpha}^T / \partial \boldsymbol{\beta}\right)\left(\partial \boldsymbol{U}_a^T / \partial \boldsymbol{\alpha}\right) + \partial \boldsymbol{U}_a^T / \partial \boldsymbol{\beta} = \boldsymbol{0}$$

and so we have for $\partial\boldsymbol{\alpha}^T/\partial\boldsymbol{\beta}$

$$\partial\boldsymbol{\alpha}^T/\partial\boldsymbol{\beta} = -\partial\boldsymbol{U}_a^T/\partial\boldsymbol{\beta}\left(\partial\boldsymbol{U}_a^T/\partial\boldsymbol{\alpha}\right)^{-1} \tag{46}$$

where again $\boldsymbol{\alpha} = \hat{\boldsymbol{\alpha}}(\boldsymbol{\beta})$. Now, substituting (46) in (45) and using (41) we obtain

$$-\boldsymbol{i}_b^* = \boldsymbol{B}^T\boldsymbol{\Omega}\boldsymbol{B} - \boldsymbol{B}^T\boldsymbol{\Omega}\boldsymbol{A}\left(\boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{B} \tag{47}$$

The matrices $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{\Omega}$ can be very large but (47) can be rewritten in a form suitable for computation. Since the columns of $\boldsymbol{A}$ are dummy variables indicating the sites and $\boldsymbol{\Omega}$ has the same block diagonal structure as $\boldsymbol{V}$ (and $\boldsymbol{R}$) we can write $\boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{A} = \text{diag}(\boldsymbol{d})$ with $\boldsymbol{d}$ the I-vector with elements $d_i = \mathbf{1}_{o_i}^T\boldsymbol{\Omega}_i\mathbf{1}_{o_i}$ and $\boldsymbol{\Omega}_i$ the i^{th} block of $\boldsymbol{\Omega}$ which can be expressed as $\boldsymbol{\Omega}_i = \text{diag}(\boldsymbol{\mu}_i)\boldsymbol{V}_i^{-1}\,\text{diag}(\boldsymbol{\mu}_i)$

Now, we can rewrite (47) as

$$\begin{aligned}-\boldsymbol{i}_b^* &= \sum_i\left(\boldsymbol{B}_i^T\boldsymbol{\Omega}_i\boldsymbol{B}_i - \frac{1}{d_i}\boldsymbol{B}_i^T\boldsymbol{\Omega}_i\boldsymbol{A}_i\boldsymbol{A}_i^T\boldsymbol{\Omega}_i\boldsymbol{B}_i\right)\\ &= \sum_i\boldsymbol{B}_i^T\left(\boldsymbol{\Omega}_i - \frac{1}{d_i}\boldsymbol{\Omega}_i\mathbf{1}_{o_i}\mathbf{1}_{o_i}^T\boldsymbol{\Omega}_i\right)\boldsymbol{B}_i,\end{aligned} \tag{48}$$

and so, the matrix $-\boldsymbol{i}_b^*$ can be built up by a summation of components for each site that do not involve very large matrices.

In summary, the algorithm alternates between updating $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ according to

$$\begin{aligned}\boldsymbol{\alpha}_i^t &= \ln\left(\boldsymbol{z}_i^T\boldsymbol{f}_i\right) - \ln\left(\boldsymbol{z}_i^T\exp(\boldsymbol{B}_i\boldsymbol{\beta}^{t-1})\right)\\ \boldsymbol{\mu}^t &= \exp(\boldsymbol{A}\boldsymbol{\alpha}^t + \boldsymbol{B}\boldsymbol{\beta}^{t-1} - \ln\boldsymbol{w})\\ \boldsymbol{\beta}^t &= \boldsymbol{\beta}^{t-1} - \left(\boldsymbol{i}_b^*\right)^{-1}\boldsymbol{U}_b^*\end{aligned} \tag{49}$$

where $\boldsymbol{z}_i^T = 1_{o_i}^T$ in case of a ML model or iteration step, and $\boldsymbol{z}_i^T = \boldsymbol{\mu}_i\boldsymbol{V}_i^{-1}$ for a GEE iteration step. For GEE steps, σ^2 and ρ are updated using the current value of $\boldsymbol{\mu}$. Equations (49) are repeated until convergence in $\boldsymbol{\beta}$ (ML) or β, ρ and σ^2 (GEE).

The asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ can be estimated by the $p_b \times p_b$ submatrix in the lower-right corner of $-\boldsymbol{i}(\boldsymbol{\theta})^{-1}$ evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. But, using the formula for the inverse of a partitioned matrix, it can be seen that this inverse equals the inverse of the right-hand side of (47) evaluated at the estimates $\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\beta}}$. So, after convergence of the algorithm (49) the matrix $-\left(\boldsymbol{i}_b^*\right)^{-1}$ provides an estimate of the covariance matrix of $\hat{\boldsymbol{\beta}}$, $\hat{\boldsymbol{\gamma}}$ (and/or $\hat{\boldsymbol{\delta}}$.

4 Model variants and extensions

4.1 Stepwise refinement

If the slope parameters (or, if covariates are present, the effects of covariates on the slope) before and after a certain changepoint do not differ significantly, one may wish to delete that changepoint in order to obtain a more parsimonious model, which has less parameters than the original model, without compromising the explanatory power. After refitting the reduced model one may again wish to delete a certain changepoint and so on. In `rtrim` a stepwise model selection procedure is implemented for this purpose. This procedure repeats the following steps:

1. Wald statistics for the difference of the parameters before and after each changepoint and their associated significance levels are calculated. If the largest significance level exceeds a certain threshold value (probability to remove, P_R, default value is 0.20) the corresponding changepoint is removed from the model.
2. For all removed changepoints except the last one, a score statistic is calculated to assess the significance of the difference in parameters before and after the changepoint. If the smallest significance level is smaller than a threshold value (probability to enter, P_E, default value is 0.15) the changepoint is added to the model.

The procedure stops if no changepoints can be either removed or added.

4.1.1 Score test

The score test mentioned above is a test for the significance of additional parameters that could be added to a model. The test can be performed without actually having to estimate the extended model that includes these additional parameters. This is especially an advantage for forward stepwise model selection procedures where the significance of a number of possible additional parameters is evaluated before adding the most significant one to the model.

The score test for testing if r-parameters among a larger set of p-parameters are significantly different from zero, is obtained as follows. Let the p-vector with parameters be partitioned as $\boldsymbol{\beta}_p = (\boldsymbol{\beta}_q, \boldsymbol{\beta}_r)$. Where $\boldsymbol{\beta}_q$ are unrestricted parameters and $\boldsymbol{\beta}_r$ are parameters restricted to be zero under the null hypothesis. To estimate $\boldsymbol{\beta}_p$ under the null hypothesis, we estimate $\boldsymbol{\beta}_q$ in a reduced model that does not contain the restricted parameters and then add zeroes for the restricted parameters, thus we obtain $\hat{\boldsymbol{\beta}}_p = (\hat{\boldsymbol{\beta}}_q, \mathbf{0}_r)$. Using $\hat{\boldsymbol{\beta}}_p$ we can evaluate the score vector and Hessian matrix under H_0 resulting in, using the notation of section 3.4, $\boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_p)$ and $\boldsymbol{i}_b^*(\hat{\boldsymbol{\beta}}_p)$, respectively.

The score statistic for testing $H_0 : \boldsymbol{\beta}_r = 0$ is then given by (see, e.g. Cox and Hinkley, 1974, Ch. 9):

$$S(\boldsymbol{\beta}_r) = \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_p)^T [-\boldsymbol{i}_b^*(\hat{\boldsymbol{\beta}}_p)]^{-1} \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_p) \tag{50}$$

But $\boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_p) = [\boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_q), \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r)] = [\mathbf{0}, \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r)]$, because of the maximisation with respect to the unrestricted parameters $\boldsymbol{\beta}_q$. Now, if we denote, for ease of notation, $[-\boldsymbol{i}_b^*(\hat{\boldsymbol{\beta}}_p)]^{-1}$ by $\boldsymbol{V}$ and we partition this matrix conformably with the partitioning $\boldsymbol{\beta}_p = (\boldsymbol{\beta}_q, \boldsymbol{\beta}_r)$, we can write the score statistic as

$$S(\boldsymbol{\beta}_r) = \begin{bmatrix} \mathbf{0} & \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r) \end{bmatrix}^T \begin{bmatrix} \boldsymbol{V}_{q,q} & \boldsymbol{V}_{q,r} \\ \boldsymbol{V}_{r,q} & \boldsymbol{V}_{r,r} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r) \end{bmatrix} = \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r)^T \boldsymbol{V}_{r,r} \boldsymbol{U}_b^*(\hat{\boldsymbol{\beta}}_r) \tag{51}$$

4.2 Reparameterization of the time effects model

Here we consider the reparameterization of the time-effects model in terms of a model with a linear trend and deviations from this linear trend for each time point. The time-effects model is given by

$$\ln \mu_{ij} = \alpha_i + \gamma_j, \tag{52}$$

with γ_j the effect for time j on the log-expected counts and $\gamma_1 \equiv 0$. This reparameterization can be expressed as

$$\ln \mu_{ij} = \alpha_i^* + \beta^* d_j + \gamma_j^*, \tag{53}$$

with $d_j = j - \bar{j}$ and $\bar{k}j$ the mean of the integers j representing the time points.

The parameter α_i^* is the intercept and the parameter β^* is the slope of the least squares regression line through the J log-expected time counts in site i and γ_j^* can be seen as the residuals of this linear fit. From regression theory we have that the 'residuals' γ_j^* sum to zero and are orthogonal to the explanatory variable, i.e.,

$$\sum_j \gamma_j^* = 0 \tag{54a}$$

and

$$\sum_j d_j \gamma_j^* = 0. \tag{54b}$$

Using these constraints we obtain the equations:

$$\ln \mu_{ij} = \alpha_i^* + \beta^* d_j + \gamma_j^* = \alpha_i + \gamma_j \tag{55}$$

$$\sum_j \ln \mu_{ij} = J\alpha_j^* = J\alpha_i + \sum_j \gamma_j \tag{56}$$

$$\sum_j d_j \ln \mu_{ij} = \beta^* \sum_j d_j^2 = \sum_j d_j \gamma_j, \tag{57}$$

where (55) is the re-parameterization equation itself and (56) and (57) are obtained by using the constraints.

From (56) we have that $\alpha_i^* = \alpha_i + \frac{1}{J}\sum_j \gamma_j$. Now, by using the equations (55) thru (57) and defining $D = \sum_j d_j^2$, we can express the parameters β^* and γ^* as functions of the parameters γ as follows:

$$\beta^* = \frac{1}{D}\sum_j d_j \gamma_j, \tag{58}$$

$$\begin{aligned}\gamma_j^* &= \alpha_i + \gamma_j - \alpha_i^* - \beta^* d_j \quad \text{(using (55))} \\ &= \alpha_i - \left(\alpha_i + \frac{1}{J}\sum_j \gamma_j\right) + \gamma_j - d_j \frac{1}{D}\sum_j d_j \gamma_j \\ &= \gamma_j - \frac{1}{J}\sum_j \gamma_j - d_j \frac{1}{D}\sum_j d_j \gamma_j. \end{aligned} \tag{59}$$

Since β^* and γ_j^* are linear functions of the parameters γ_j they can be expressed in matrix notation by

$$\begin{pmatrix} \beta^* \\ \boldsymbol{\gamma}^* \end{pmatrix} = \boldsymbol{T}\boldsymbol{\gamma}, \tag{60}$$

with $\boldsymbol{\gamma}^* = (\gamma_1^*, \dots, \gamma_J^*)$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_J)$ and $\boldsymbol{T}$ the $(J+1) \times J$ transformation matrix that transforms $\boldsymbol{\gamma}$ to $(\beta^*, \boldsymbol{\gamma}^*)$. From (58) and (59) it follows that the elements of $\boldsymbol{T}$ are given by:

$$\begin{aligned}
\boldsymbol{T}_{(1,j)} &= \frac{d_j}{D} && \text{(for } i = 1; j = 1, \dots, J\text{)} \\
\boldsymbol{T}_{(i,j)} &= 1 - \frac{1}{j} - \frac{1}{D} d_{i-1} d_j && \text{(for } i = 2, \dots, J+1; j = 1, \dots, J; i-1 = j\text{)} \\
\boldsymbol{T}_{(i,j)} &= \quad - \frac{1}{j} - \frac{1}{D} d_{i-1} d_j && \text{(for } i = 2, \dots, J+1; j = 1, \dots, J; i-1 \neq j\text{)}
\end{aligned}$$

The covariance matrix of the transformed parameter vector can now be obtained from the covariance matrix of $\boldsymbol{\gamma}$ as

$$\text{cov}\begin{pmatrix} \beta^* \\ \boldsymbol{\gamma}^* \end{pmatrix} = \text{cov}(\boldsymbol{T}\boldsymbol{\gamma}) = \boldsymbol{T}\,\text{cov}(\boldsymbol{\gamma})\boldsymbol{T}^T. \tag{61}$$

4.2.1 Wald-test for deviations from linear trend

To test for the significance of the deviations of the linear trend, we can test the hypothesis $H_0 : \boldsymbol{\gamma}^* = \mathbf{0}$. To test this H_0 we must take into account that two of the γ^*-parameters are redundant in the sense that any subset of 2 of the γ^*-parameters can be obtained from the remaining $J-2$ ones by using the two linear constraint equations (54). In particular, the values of any subset of 2 parameters are zero if the remaining $J-2$ ones are. Therefore, testing $\boldsymbol{\gamma}^* = \mathbf{0}$ is equivalent to testing $\boldsymbol{\gamma}_{J-2}^* = \mathbf{0}$, with $\boldsymbol{\gamma}_{J-2}^*$ a vector consisting of some subset of $J-2$ of the elements of $\boldsymbol{\gamma}^*$. The Wald-statistic for $H_0 : \boldsymbol{\gamma}_{J-2}^* = \mathbf{0}$ is given by

$$W_{\boldsymbol{\gamma}_{J-2}^*} = (\boldsymbol{\gamma}_{J-2}^*)^T \,\text{var}\left(\boldsymbol{\gamma}_{J-2}^*\right)^{-1} \boldsymbol{\gamma}_{J-2}^*, \tag{62}$$

which is independent of the choice of the subset of $J-2$ γ^*-parameters. This statistic is asymptotically χ^2 distributed with $J-2$ degrees of freedom.

Alternatively, we could retain the complete $\boldsymbol{\gamma}^*$-vector and the corresponding covariance matrix to define a Wald-statistic. In that case, a generalized inverse must be used since the covariance-matrix of $\boldsymbol{\gamma}^*$ is a singular matrix, with rank $J-2$ (See e.g. Harville, 1997, Chapter 9). This approach would, however, lead to the same results as using (62).

5 Uncertainty analysis

The most important parameters produced by `rtrim` are the time-totals and, especially, the indices that are derived from them. In this subsection we describe how the covariance matrix of the time-totals is estimated and how the covariance matrix of the indices is derived from that matrix. For the calculation of these covariance matrices we must distinguish between model-based and imputed time-totals and indices (see, subsection 2.1). In subsections 5.2.1 and 5.2.2 we describe the covariance estimator for the model-based and imputed time-totals, respectively, and in subsection 5.3 we show how the variance of indices can be derived from the variance of time-totals.

5.1 Intermezzo: Standard error of multiplicative parameters and the delta-method

The multiplicative parameters are simple transformations of the additive parameters. If we let θ denote an additive scalar parameter, then the corresponding multiplicative scalar parameter, t say, can be expressed as a function of the additive parameter by $t = f(\theta)$, with $f = \exp(.)$.

The variance of the multiplicative parameters can be approximated by the use of the delta method (see, e.g. Agresti, 1990, ch. 12 or Särndal et al., 1992, ch. 5). This method is based on approximating the function $f(\hat{\theta})$ by the first two terms of the Taylor series expansion around the true value θ:

$$f(\hat{\theta}) = f(\theta) + f'(\theta)(\hat{\theta} - \theta),$$

with $f'(\theta)$ the derivative of f w.r.t. θ. The variance of this approximation is

$$\text{var}(\hat{t}) = \text{var}\left(f(\hat{\theta})\right) = \left(f'(\theta)\right)^2 \text{var}(\hat{\theta}). \tag{63}$$

For the function $f = \exp(\cdot)$, that transforms the additive parameters into the multiplicative ones, this variance approximation leads to the variance estimator

$$\text{var}(\hat{t}) = \hat{t}^2 \, \text{var}(\hat{\theta}), \tag{64}$$

so that the standard error of a multiplicative parameter can simply be estimated by the standard error estimate of the additive parameter times the estimated value of the multiplicative parameter.

There is a straightforward generalisation of the Taylor-series approximation, and corresponding variance estimator, for a (vector- or scalar-valued) function of a vector-valued random variable. In particular, if $\boldsymbol{t}$ is a function $f(\boldsymbol{\theta})$ of $\boldsymbol{\theta}$, with $\boldsymbol{\theta}$ a vector and $\boldsymbol{t}$ either a vector or a scalar, the first two terms of the Taylor-series expansion are

$$f(\hat{\boldsymbol{\theta}}) = f(\boldsymbol{\theta}) + \boldsymbol{D}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \tag{65}$$

with $\boldsymbol{D} = \partial f(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ and the covariance matrix of $\boldsymbol{t}$ can be expressed as

$$\text{cov}(\boldsymbol{t}) = \boldsymbol{D} \, \text{cov}(\boldsymbol{\theta}) \boldsymbol{D}^T, \tag{66}$$

with $\text{cov}(\boldsymbol{\theta})$ the covariance matrix of $\boldsymbol{\theta}$, a result which will be used repeatedly in the sections below.

5.2 Standard errors of time-totals

5.2.1 Standard errors of model based time-totals

Starting with model based time totals, these are defined as the total estimated counts for a given year j aggregated over all sites i:

$$\hat{t}_j = \sum_{i(m)} \hat{\mu}_{ij(m)} \tag{67}$$

or, vectorizing,

$$\hat{\boldsymbol{t}} = \boldsymbol{C}\hat{\boldsymbol{\mu}} \tag{68}$$

where $\hat{\boldsymbol{t}}$ is a J-vector, and $J \times IJ$ matrix $\boldsymbol{C}$ is defined as $\boldsymbol{C} = \left(\boldsymbol{I}_J, \boldsymbol{I}_J, \ldots, \boldsymbol{I}_J\right)$, with $\boldsymbol{I}_J$ an $J \times J$ identity matrix, and the number of identity matrices equal to the number of sites I.

The covariance matrix of the estimated time-totals can then be expressed as

$$\begin{aligned} \text{var}(\hat{\boldsymbol{t}}) &= \boldsymbol{C}\,\text{var}(\hat{\boldsymbol{\mu}})\boldsymbol{C}^T = \boldsymbol{C}\boldsymbol{D}\,\text{var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}^T\boldsymbol{C}^T \\ &= \boldsymbol{C}\,\text{diag}(\hat{\boldsymbol{\mu}})\boldsymbol{X}\,\text{var}(\hat{\boldsymbol{\theta}})\boldsymbol{X}^T\,\text{diag}(\hat{\boldsymbol{\mu}})\boldsymbol{C}^T \end{aligned} \tag{69}$$

where we have used the usual Taylor-series variance approximation $\text{var}(\hat{\boldsymbol{\mu}}) = \boldsymbol{D}\,\text{var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}^T$, with $\boldsymbol{D} = \partial\boldsymbol{\mu}/\partial\boldsymbol{\theta}^T$ (see section 5.1 and the paragraph following Eqn (30)).

To compute the standard errors of the model based time-totals according to (69) we need the covariance matrix of the complete estimated parameter vector $\boldsymbol{\theta}$. This matrix is not easy to compute because it requires inversion of a very large matrix, as pointed out in section 3.2. An alternative formula for $\text{var}(\hat{\boldsymbol{t}})$ that is suitable for computation will be derived in this subsection.

Consider the partitioned form (41) of $\boldsymbol{i}(\boldsymbol{\theta})$, which can be re-expressed as

$$\boldsymbol{i}(\boldsymbol{\theta}) = \left[\begin{array}{c|c} \boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{A} & \boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{B} \\ \hline \boldsymbol{B}^T\boldsymbol{\Omega}\boldsymbol{A} & \boldsymbol{B}^T\boldsymbol{\Omega}\boldsymbol{B} \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{i}_{aa} & \boldsymbol{i}_{ab} \\ \hline \boldsymbol{i}_{ab}^T & \boldsymbol{i}_{bb} \end{array}\right] \tag{70}$$

with $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\mu})\boldsymbol{V}^{-1}\,\text{diag}(\boldsymbol{\mu})$.

The inverse of this partitioned matrix can be expressed as (Rao, 1973, page 33)

$$\boldsymbol{i}(\boldsymbol{\theta})^{-1} = \left[\begin{array}{c|c} \boldsymbol{i}_{aa}^{-1} + \boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{F}^T & -\boldsymbol{F}\boldsymbol{E}^{-1} \\ \hline -\boldsymbol{E}^{-1}\boldsymbol{F}^T & \boldsymbol{E}^{-1} \end{array}\right] = \left[\begin{array}{c|c} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \hline \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} \end{array}\right] \tag{71}$$

where

$$\boldsymbol{i}_{aa} = \boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{A} = \text{diag}(\boldsymbol{d}),$$
$$\boldsymbol{E} = \boldsymbol{i}_{bb} - \boldsymbol{i}_{ab}^T\boldsymbol{i}_{aa}^{-1}\boldsymbol{i}_{ab} = \boldsymbol{i}_{\beta}^*$$

and

$$\boldsymbol{F} = \boldsymbol{i}_{aa}^{-1}\boldsymbol{i}_{ab} = \text{diag}(\boldsymbol{d})^{-1}\boldsymbol{A}^T\boldsymbol{\Omega}\boldsymbol{B} = \begin{bmatrix} d_1^{-1}\boldsymbol{w}_1^T\boldsymbol{B}_1 \\ \vdots \\ d_I^{-1}\boldsymbol{w}_I^T\boldsymbol{B}_I \end{bmatrix},$$

with

$$\boldsymbol{w}_i = \boldsymbol{1}_{O_i}^T\boldsymbol{\Omega}_i = \sum_j (\boldsymbol{\mu}_i^{\frac{1}{2}})_j (\boldsymbol{\mu}_i^{\frac{1}{2}})_k (\boldsymbol{R}_i)_{jk}$$

and

$$d_i = \mathbf{1}_{O_i}^T \boldsymbol{w}_i$$

From this representation we see that we only need the inverse of $\boldsymbol{E}$ (which is already produced by the algorithm) and the inverse of $\boldsymbol{i}_{aa}$ (which is a diagonal matrix) in order to calculate the inverse of $\boldsymbol{i}(\boldsymbol{\theta})$.

The covariance matrix of the estimated time-totals can now be expressed as

$$\begin{aligned}
\text{var}(\hat{\boldsymbol{t}}) &= \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})\boldsymbol{X}\,\text{var}(\hat{\boldsymbol{\theta}})\boldsymbol{X}^T\,\text{diag}(\boldsymbol{\mu})\boldsymbol{C}^T \\
&= \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})\left[\boldsymbol{A}\boldsymbol{\Phi}_{11}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{\Phi}_{21}\boldsymbol{A}^T + \boldsymbol{A}\boldsymbol{\Phi}_{12}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Phi}_{22}\boldsymbol{B}^T\right]\text{diag}(\boldsymbol{\mu})\boldsymbol{C}^T \\
&= \boldsymbol{G}\boldsymbol{\Phi}_{11}\boldsymbol{G}^T + \boldsymbol{H}\boldsymbol{\Phi}_{21}\boldsymbol{G}^T + \boldsymbol{G}\boldsymbol{\Phi}_{12}\boldsymbol{H}^T + \boldsymbol{H}\boldsymbol{\Phi}_{22}\boldsymbol{H}^T \\
&= \boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T + \boldsymbol{G}\boldsymbol{F}\boldsymbol{E}^{-1}(\boldsymbol{G}\boldsymbol{F})^T - \boldsymbol{H}\boldsymbol{E}^{-1}(\boldsymbol{G}\boldsymbol{F})^T - \boldsymbol{G}\boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{H}^T + \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{H}^T \\
&= \boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T + (\boldsymbol{G}\boldsymbol{F} - \boldsymbol{H})\boldsymbol{E}^{-1}(\boldsymbol{G}\boldsymbol{F} - \boldsymbol{H})^T
\end{aligned} \tag{72}$$

with $\boldsymbol{G} = \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})\boldsymbol{A}$ and $\boldsymbol{H} = \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})\boldsymbol{B}$.

To compute the covariance matrix (72) the following expressions for the elements of the $J \times I$ matrix $\boldsymbol{G}$, the $J \times p_b$ matrix $\boldsymbol{G}\boldsymbol{H}$, the $J \times p_b$ matrix $\boldsymbol{H}$ and the $J \times J$ matrix $\boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T$ are used:

$$\boldsymbol{G}_{ji} = \mu_{ij}, \tag{73}$$

$$(\boldsymbol{G}\boldsymbol{F})_{jk} = \sum_i \mu_{ij}\boldsymbol{F}_{ik}, \tag{74}$$

$$\boldsymbol{H}_{jk} = \sum_i (\boldsymbol{B}_i)_{jk}\mu_{ij} \tag{75}$$

$$(\boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T)_{jk} = \sum_i \mu_{ij}\mu_{ik}d_i^{-1}. \tag{76}$$

So, the matrices $\boldsymbol{G}\boldsymbol{F} - \boldsymbol{H}$ and $\boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T$ can be obtained by a summation over sites.

In case of monthly data, the expressions for $\boldsymbol{G}$ and $\boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T$ changes to

$$\begin{aligned}
\boldsymbol{G}_{ji} &= \sum_m \mu_{ijm} \\
(\boldsymbol{G}\boldsymbol{F})_{jk} &= \sum_i \boldsymbol{G}_{ji}\boldsymbol{F}_{ik} \\
\boldsymbol{H}_{jk} &= \sum_i \sum_m (\boldsymbol{B}_{im})_{jk}\mu_{ijm} \\
(\boldsymbol{G}\text{diag}(\boldsymbol{d})^{-1}\boldsymbol{G}^T)_{jk} &= \sum_i \boldsymbol{G}_{ji}\boldsymbol{G}_{ik}d_i^{-1}
\end{aligned}$$

where $\boldsymbol{B}_{im}$ the subblock of $\boldsymbol{B}_i$ representing the β parameters corresponding to month m. For $m = 1$ these are rows $1, \ldots, J$, for $m = 2$ rows $(J+1), \ldots, 2J$, etc.

5.2.2 Standard error of imputed time-totals

The J-vector with imputed time-totals can be written as

$$\tilde{\boldsymbol{t}} = \boldsymbol{C}\boldsymbol{f}^+ = \boldsymbol{C}_o\boldsymbol{f} + \boldsymbol{C}_x\hat{\boldsymbol{\mu}}_x, \tag{77}$$

where $\boldsymbol{f}^+$ denotes the vector containing the imputed counts with elements given by

$$f^+_{ij(m)} = \begin{cases} f_{ij(m)} & \text{if observed} \\ \mu_{ij(m)} & \text{otherwise} \end{cases}$$

The vector $\boldsymbol{f}^+$ is split up in two parts: a vector $\boldsymbol{f}$ containing the observed elements and a vector $\hat{\boldsymbol{\mu}}_x$ with the estimated values for the missing observations. The matrix $\boldsymbol{C}$ is split up accordingly, into a matrix $\boldsymbol{C}_o$ containing the columns of $\boldsymbol{C}$ corresponding to the observed elements in $\boldsymbol{f}^+$ and a matrix $\boldsymbol{C}_x$ containing the columns of $\boldsymbol{C}$ corresponding to the missing elements in $\boldsymbol{f}^+$.

Now, the covariance matrix of $\tilde{\boldsymbol{t}}$ can be written as the sum of three $J \times J$ matrices:

$$\text{var}(\tilde{\boldsymbol{t}}) = \boldsymbol{C}_o \,\text{var}(\boldsymbol{f})\boldsymbol{C}_o^T + 2\boldsymbol{C}_x \,\text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f})\boldsymbol{C}_o^T + \boldsymbol{C}_x \,\text{var}(\hat{\boldsymbol{\mu}}_x)\boldsymbol{C}_x^T. \tag{78}$$

where we have used that $\boldsymbol{C}_o \,\text{cov}(\boldsymbol{f}, \hat{\boldsymbol{\mu}}_x)\boldsymbol{C}_x^T = \boldsymbol{C}_x \,\text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f})\boldsymbol{C}_o^T$.

To evaluate (78) we need estimates of the three covariance matrices $\text{var}(\boldsymbol{f})$, $\text{var}(\hat{\boldsymbol{\mu}}_x)$ and $\text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f})$. Using previous results, $\text{var}(\boldsymbol{f})$ and $\text{var}(\hat{\boldsymbol{\mu}})$ are relatively easy to obtain but $\text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f})$ needs some further linear approximations.

To obtain the covariance between the observed counts $\boldsymbol{f}$ and the estimated missing counts $\hat{\boldsymbol{\mu}}_x$ we first express the estimated expected counts as a function of the estimated parameters $\hat{\boldsymbol{\theta}}$ by the Taylor-series approximation according to (65):

$$\hat{\boldsymbol{\mu}}_x \approx \boldsymbol{\mu}_x + \boldsymbol{D}_x(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \tag{79}$$

since $\boldsymbol{D}_x = \partial \mu_x / \partial \boldsymbol{\theta}$. Next we use a Taylor-series approximation of the GEE estimating equation (30) to express the parameter estimate $\hat{\boldsymbol{\theta}}$ as a function of the observed counts, leading to

$$\begin{aligned} \boldsymbol{U}(\hat{\boldsymbol{\theta}}) &= \boldsymbol{D}_o^T\boldsymbol{V}^{-1}(\boldsymbol{f} - \hat{\boldsymbol{\mu}}_o) \\ &\approx \boldsymbol{U}(\boldsymbol{\theta}) + (\partial \boldsymbol{U}(\boldsymbol{\theta})/\partial\boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \boldsymbol{D}_o^T\boldsymbol{V}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}_o) - \boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}), \end{aligned}$$

and hence,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = (\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o)^{-1}\boldsymbol{D}_o^T\boldsymbol{V}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}_o),$$

which, after substituting in (79), results in an expression for $\hat{\boldsymbol{\mu}}_x$ as a function of $\boldsymbol{f}$:

$$\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x = \boldsymbol{D}_x(\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o)^{-1}\boldsymbol{D}_o^T\boldsymbol{V}^{-1}(\boldsymbol{f} - \boldsymbol{\mu}_o).$$

Also,

$$\text{var}(\hat{\boldsymbol{\theta}}) = (\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o)^{-1}.$$

Using these expression we obtain for $\text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f})$

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\mu}}_x, \boldsymbol{f}) &= E(\hat{\boldsymbol{\mu}}_x - \boldsymbol{\mu}_x)(\boldsymbol{f} - \boldsymbol{\mu}_o)^T \\ &= \boldsymbol{D}_x(\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o)^{-1}\boldsymbol{D}_o^T\boldsymbol{V}^{-1}E(\boldsymbol{f} - \boldsymbol{\mu}_o)(\boldsymbol{f} - \boldsymbol{\mu}_o)^T \\ &= \boldsymbol{D}_x(\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\boldsymbol{D}_o)^{-1}\boldsymbol{D}_o^T\boldsymbol{V}^{-1}\,\text{var}(\boldsymbol{f}) \\ &= \boldsymbol{D}_x\,\text{var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}_o^T \qquad (\text{using } \boldsymbol{V} = \text{var}(\boldsymbol{f})). \end{aligned} \tag{80}$$

Now, by using $\boldsymbol{C}_x \text{ var}(\hat{\boldsymbol{\mu}})\boldsymbol{C}_x^T = \boldsymbol{C}_x\boldsymbol{D}_x \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}_x^T\boldsymbol{C}_x^T$ (analogous to (69)) and by substituting (80) in (78), we have

$$\begin{aligned}\text{var}(\tilde{\boldsymbol{t}}) &= \boldsymbol{C}_o \text{ var}(\boldsymbol{f})\boldsymbol{C}_o^T + 2\boldsymbol{C}_x\boldsymbol{D}_x \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}_o^T\boldsymbol{C}_o^T + \boldsymbol{C}_x\boldsymbol{D}_x \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}_x^T\boldsymbol{C}_x^T, \\ &= \boldsymbol{C}_o \text{ var}(\boldsymbol{f})\boldsymbol{C}_o^T + \boldsymbol{C}\boldsymbol{D} \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}^T\boldsymbol{C}^T - \boldsymbol{C}_o\boldsymbol{D}_o \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}_o^T\boldsymbol{C}_o^T. \end{aligned} \quad (81)$$

The matrix $\boldsymbol{C}_o \text{ var}(\boldsymbol{f})\boldsymbol{C}_o^T$ can be estimated from the data by using $\text{var}(\boldsymbol{f})$ as decribed in section 3.3. This estimate depends on the assumptions of possible serial correlation and overdispersion. If serial correlation is assumed to be present, the observed counts within sites are correlated but remain independent across sites, resulting in a block-diagonal structure with blocks corresponding to the sites.

The matrix $\boldsymbol{C}\boldsymbol{D} \text{ var}(\hat{\boldsymbol{\theta}})\boldsymbol{D}^T\boldsymbol{C}^T$ is the covariance matrix of the model based time-totals, the computation of which is described in section 5.2.1. The matrix $\boldsymbol{C}_o\boldsymbol{D}_o var(\hat{\boldsymbol{\theta}})\boldsymbol{D}_o^T\boldsymbol{C}_o^T$ can be calculated similarly, by restricting all calculations to the observed counts only.

5.2.3 Standard errors with external covariance matrix of counts

The estimation procedures and standard error estimates in `rtrim` normally use an estimate of the covariance matrix of the observed counts, based on the user specified options for serial correlation and overdispersion. It is, however, also possible to use a covariance matrix that is completely specified by the user. In this case, the parameters of the models will be estimated by maximum likelihood, that is, using a covariance matrix based on the assumption of independent Poisson distributions for the counts (no serial correlation, variance equal to the expected value). Although this assumption will normally not be in line with the user specified covariance matrix, the parameter estimates remain consistent and the effects on point estimates of using the ”wrong” covariance matrix are usually small. This does, however, not hold for the effects on variances and standard errors and for these a correction is necessary that takes the user specified covariance matrix into account. This approach is described by e.g. Royal (1986) and White (1980). The resulting corrected covariance estimator is often called the “sandwich” estimator. This sandwich estimator is applied in `rtrim` for estimating the covariance of time-totals when a user specified covariance matrix is used and will be described below.

The sandwich covariance estimator is given by

$$\text{var}(\hat{\boldsymbol{\theta}}) = \boldsymbol{i}(\boldsymbol{\theta})^{-1}\boldsymbol{S}\,\boldsymbol{i}(\boldsymbol{\theta})^{-1} \quad (82)$$

With $\boldsymbol{i}(\boldsymbol{\theta})^{-1}$ the inverse of the information matrix of the likelihood (compare (32)). This matrix is the covariance matrix of the parameter estimate $\hat{\boldsymbol{\theta}}$ if the model assumptions underlying the ML-procedure are satisfied (independent Poisson counts). The matrix $\boldsymbol{i}(\boldsymbol{\theta})$ in (82) is a special case of the corresponding matrix in the GEE-estimation procedure of section 3.2, obtained by setting the covariance matrix of the observed counts ($\boldsymbol{V}$) equal to $\text{diag}(\boldsymbol{\mu})$ in accordance with the independent Poisson assumption.

The matrix $\boldsymbol{S}$ is the outer product of the score vector (derivative of the log-likelihood), which in this case is:

$$\begin{aligned}\boldsymbol{S} &= \text{E}\left[\boldsymbol{X}^T(\boldsymbol{f}-\boldsymbol{\mu})(\boldsymbol{f}-\boldsymbol{\mu})^T\boldsymbol{X}\right] \\ &= \boldsymbol{X}^T \text{ var}(\boldsymbol{f})\boldsymbol{X}, \end{aligned} \quad (83)$$

with $\boldsymbol{f}$ the observed counts and $\boldsymbol{\mu}$ the expected counts corresponding to these observed counts. The covariance matrix of the observed counts var$(\boldsymbol{f})$ is a block diagonal matrix with elements specified by the user. Using this $\boldsymbol{S}$ we write the covariance matrix of $\hat{\boldsymbol{t}}$ as:

$$\begin{aligned}\text{var}(\hat{\boldsymbol{t}}) &= \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})\boldsymbol{X}\boldsymbol{i}(\boldsymbol{\theta})^{-1}\boldsymbol{X}^T\,\text{var}(\boldsymbol{f})\boldsymbol{X}\boldsymbol{i}(\boldsymbol{\theta})^{-1}\boldsymbol{X}^T\,\text{diag}(\boldsymbol{\mu})\boldsymbol{C}^T \\ &= \boldsymbol{P}\,\text{var}(\boldsymbol{f})\boldsymbol{P}^T, \text{ say.}\end{aligned} \tag{84}$$

Using the partitioning (70) of $\boldsymbol{i}(\boldsymbol{\theta})^{-1}$ and $\boldsymbol{X} = [\boldsymbol{A}, \boldsymbol{B}]$ we can write for $\boldsymbol{P}$:

$$\begin{aligned}\boldsymbol{P} &= \boldsymbol{C}\,\text{diag}(\boldsymbol{\mu})[\boldsymbol{A}\boldsymbol{\Phi}_{11}\boldsymbol{A}^T + \boldsymbol{B}\boldsymbol{\Phi}_{21}\boldsymbol{A}^T + \boldsymbol{A}\boldsymbol{\Phi}_{12}\boldsymbol{B}^T + \boldsymbol{B}\boldsymbol{\Phi}_{22}\boldsymbol{B}^T \\ &= \boldsymbol{G}\boldsymbol{\Phi}_{11}\boldsymbol{A}^T + \boldsymbol{H}\boldsymbol{\Phi}_{21}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{\Phi}_{12}\boldsymbol{B}^T + \boldsymbol{H}\boldsymbol{\Phi}_{22}\boldsymbol{B}^T \\ &= \boldsymbol{G}\text{diag}(\boldsymbol{\mu}_+)^{-1}\boldsymbol{A}^T + \boldsymbol{G}\boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{F}^T\boldsymbol{A}^T - \boldsymbol{G}\boldsymbol{F}\boldsymbol{E}^{-1}\boldsymbol{B}^T - \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{F}^T\boldsymbol{A}^T + \boldsymbol{H}\boldsymbol{E}^{-1}\boldsymbol{B}^T \\ &= \boldsymbol{G}\text{diag}(\boldsymbol{\mu}_+)^{-1}\boldsymbol{A}^T + (\boldsymbol{G}\boldsymbol{F} - \boldsymbol{H})\boldsymbol{E}^{-1}(\boldsymbol{F}^T\boldsymbol{A}^T - \boldsymbol{B}^T),\end{aligned} \tag{85}$$

where $\boldsymbol{F}$, $\boldsymbol{G}$, $\boldsymbol{H}$ and $\boldsymbol{E}$ as defined in (71) and (72). The matrices $\boldsymbol{F}$ and $\boldsymbol{E}$ are somewhat simpler here because, in the absence of overdispersion and serial correlation, they are based on the simplified form of $\boldsymbol{i}(\boldsymbol{\theta})$ obtained by setting $\boldsymbol{\Omega}$ equal to diag$(\boldsymbol{\mu})$. The matrix diag$(\boldsymbol{\mu}_+)$ is a diagonal matrix with the sum, over the years, of the expected counts of each site on the diagonal.

The first component of the sum $\boldsymbol{P}$ (85), can be partitioned according to the sites as,

$$\begin{aligned}\boldsymbol{G}\text{diag}(\boldsymbol{\mu}_+)^{-1}\boldsymbol{A}^T &= \frac{1}{\mu_{+1}}\boldsymbol{\mu}_1\mathbf{1}_{\text{obs}_1}^T, \ldots, \frac{1}{\mu_{+i}}\boldsymbol{\mu}_i\mathbf{1}_{\text{obs}_i}^T, \ldots, \frac{1}{\mu_{+I}}\boldsymbol{\mu}_i\mathbf{1}_{\text{obs}_I}^T \\ &= [\boldsymbol{Q}_1 \vdots, \ldots, \vdots\, \boldsymbol{Q}_i \vdots, \ldots, \vdots\, \boldsymbol{Q}_I], \text{ say,}\end{aligned} \tag{86}$$

with $\mathbf{1}_{\text{obs}_i}$ a vector of ones with length equal to obs$_i$, the number of observed values in site i, thus the length of $\boldsymbol{f}_i$. The site specific matrices $\boldsymbol{Q}_i$ have obs$_i$ identical columns, each equal to μ_{ij}/μ_{i+}.

The matrix $(\boldsymbol{G}\boldsymbol{F} - \boldsymbol{H})$ is a constant for all sites because it is obtained by a summation over sites, analogous to the calculation of this matrix in the case of variance of the model based indices in section 5.2.1.

The matrix $\boldsymbol{E}^{-1}$ is also constant over sites. It is in this case the estimated covariance matrix of the β-parameters according to the ML-method with which these parameters are estimated. This matrix is calculated as part of the ML-procedure.

5.3 Standard error of indices

Estimated time-totals can be model-based or imputed, and each of these is based on a model that can have been estimated the ML or GEE method. The covariance matrix of the time-totals will differ among definitions and methods, but in all cases, the indices are the same functions of the time-totals and the covariance matrix of the indices is the same function of the covariance matrix of the time-totals, irrespective of the definition or estimation procedure used.

The index for time-point j with respect to some reference time point b, (τ_j say) can be expressed as a function of the time-totals for time-points j and b:

$$\tau_j = t_j/t_b. \tag{87}$$

To obtain the variances of the estimated indices as functions of the variances and covariance of the time-totals t_j and t_b, we apply the delta-method outlined in section 5.1. For this we need the vector with derivatives of τ_j w.r.t. t_j and t_b, given by

$$\boldsymbol{d} = \begin{pmatrix} -t_j t_b^{-2} \\ t_b^{-1} \end{pmatrix} \tag{88}$$

The variance of an index τ_j can now be obtained, by applying (66), as

$$\text{var}(\tau_j) = \boldsymbol{d}^T \boldsymbol{V}_{t_b,t_j} \boldsymbol{d}, \tag{89}$$

with $\boldsymbol{V}_{t_b,t_j}$ the covariance matrix of t_b and t_j corresponding to the definition of the totals (model-based or imputed) and estimation method used.

Note that the index for time-point b (the base-time) is, by definition and for all data sets, equal to 1. So, for $j = b$ we must have that $\text{var}(\tau_j) = 0$. Indeed, by substituting t_b for t_j in (89) and setting $\text{var}(t_j) = \text{cov}(t_j, t_b) = \text{var}(t_b)$ we obtain zero for the r.h.s. of (89).

5.3.1 Using multi-year reference periods for indexing

Sometimes, interannual variability of observations is high, such that the computed index values are highly contingent on the counts in the base year. In these cases, it may be more robust to use a longer time period as reference instead of a single year. `rtrim` facilitates this extended indexing by allowing for multiple, say n, base years, say $b_1, \ldots, b_n$ In this case, the expression (87) for τ_j changes to

$$\tau_j = t_j / \bar{t_b} \tag{90}$$

where $\bar{t_b} = (t_{b_1} + \ldots + t_{b_n})/n$ is the average time total for the base years.

The partial derivatives of τ_j are now given by the vector of length $n + 1$

$$\boldsymbol{d} = \begin{pmatrix} -n t_j (\sum_n t_{b_i})^{-2} \\ \vdots \\ -n t_j (\sum_n t_{b_i})^{-2} \\ n (\sum_n t_{b_i})^{-1} \end{pmatrix} \tag{91}$$

where the repeated elements make up the first n elements of $\boldsymbol{d}$. Note that (87) and (88) are special cases of (90) and (91).

5.4 Standard error of overall slope

In section 2.5 we defined as a summary measure for the overall trend the slope of the regression line, estimated by ols, through the estimated log time-totals (model (15)). For this model (see, (16))

$$\hat{\boldsymbol{\beta}} = (\alpha, \beta_+)^T = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{y},$$

with $\boldsymbol{y}$ the vector with as elements the log-expected total counts, $\ln \mu_{+j}$. The covariance matrix of $\hat{\boldsymbol{\beta}}$ is a function of the covariance matrix $\boldsymbol{V}(\boldsymbol{y})$ of $\boldsymbol{y}$ and is given by

$$\boldsymbol{V}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{V}(\boldsymbol{y}) \boldsymbol{X} (\boldsymbol{X}^T \boldsymbol{X})^{-1}, \tag{92}$$

and the variance of the estimated overall slope parameter $\hat{\beta}_+$ is the lower right element of this matrix.

The covariance matrix $V(\boldsymbol{y})$ in (92) will depend on the model used and the specification of the covarance matrix of the observed counts, e.g., the settings of the options for serial correlation and overdispersion.

References

Amemiya, T, 1985, *Advanced Econometrics*, Blackwell, London.

Agresti, A., 1990, *Categorical Data Analysis*, Wiley, New York.

Bogaart, P. W., M. van der Loo, and J. Pannekoek, 2016. *rtrim: Trends and Indices for Monitoring Data*, R package version 1.0.1 `https://CRAN.R-project.org/package=rtrim`

ter Braak, C.J.F., A.J. van Strien, R. Meijer, and T.J. Verstrael, 1994, *Analysis of monitoring data with many missing values: which method?* In: *Proceedings Bird Numbers 1992*.

Cox, D.R. and D.V. Hinkley, 1974, *Theoretical Statistics*, Chapman and Hall, London).

Fienberg, S.E., 1977, *The analysis of cross-classified categorical data*, MIT press (Cambridge, Massachusetts).

Harville, D. A., 1997, *Matrix Algebra from a Statistician's Perspective*, Springer-Verlag, New York.

Liang, K. Y. and S. L. Zeger, 1986, Longitudinal data analysis using generalized linear models, *Biometrika*, 73, 13–22.

McCullagh, P. and J. A. Nelder, 1989, *Generalized Linear Models, 2nd edition*, Chapman and Hall, London.

Rao, C. R., 1973, *Linear Statistical Inference and Its Applications, Second edition*, Wiley, New York.

Royall, R. M., 1986, Model robust confidence intervals using maximum likelihood estimators. *International Statistical Review*, *54*, 221–226.

Särndal, C.-E., B. Swensson and J. Wretman, 1992, *Model Assisted Survey Sampling*, Springer-Verlag, New York.

van Strien, A.J., J. Pannekoek, E.J.M. Hagemeijer and T.J. Verstraal, 1995, A loglinear Poisson regression method to analyze bird monitoring data, *Paper presented at: Bird Numbers Conference 1995.*

Thomas, L. 1996, Monitoring long-term population change: why are there so many analysis methods?, *Ecology*.

Underhill, L. G. and R. P. Prŷs-Jones, 1994, Index numbers for waterbird populations. I. Review and methodology, *Journal of Applied Ecology*, *31*, 463–480.

Weinreich, J.A. and J.H. Oude Voshaar, 1992, Population trends of bats hibernating in marl caves in the Netherlands (1943–1987), *Myotis*, *30*, 75–84.

White, H., 1980, A heteroscedsticity-consistent covariance matrix estimator and a direct test for heteroscedasticity, *Econometrica*, *41*, 733–750.

Zeger, S. L. and K. Y. Liang, 1986, Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, 42, 121-130.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information