**Requirement specification for recoding of A-MOMO in R**

Principally this is “just” a recoding of the current Stata A-MOMO in R. Thereby, making it more applicable e.g. not demanding Stata-license.

The grouping used here is age groups, but A-MOMO can handle groups in general i.e. if more information, like sex, is include in the input data, then grouping may depend on this too. Other possibilities could be area e.g. county.

However, a few minor improvements should be implemented:

* Parameters should all be in one or more separate parameter-file(s)  
  Now, they are in different placed in the Stata-do-files
* Delay-adjusted number of deaths (cnb) must always be greater than or equal to the registered number of deaths (nb)  
  Now, it may happen that nb>cnb
* EuroMOMO demands certain age groups. These groups should be obligatory – and only these should be included in the file to be send to EuroMOMO.  
  Other groups may be defined for local use e.g. sex or regions
* Delay-adjustment by group  
  Now, it is the same period for all groups
* Uncertainty (variance) from delay-adjustment should be included in the variance of the baseline model – and z-score calculations. At least optionally.
* Baseline always over the preceding 5 seasons, not including the present  
  By season is understood winter-season (week 40 to week 20) and summer-season (week 21 to week 39)
* Optional if seasonality (sine curve) and trend manually or automatically (testing out) should be included
* Model evaluation should be an option  
  Now, this is run every time
* Flexible selection of periods to be included or excluded in the baseline  
  Now, it is fixed to weeks in spring and autumn, but it should be possible to specify these periods more freely and varying over time – maybe there still should be an option for fixed spring and autumn?

Further, the R-code should be an R-package, with whatever this demands.

**Input**

Data

File with one record/line/row for each death, with Date of Death (DoD), Date of Registration (DoR) and age at death (age). Further, other variables may be included, to be able to make more refined groups  
- format? – csv (, or ;), Excel, others

- Work/bank holidays – for delay-adjustment

Parameters

Parameters should be in one or more separate files.

See appendix A – to be completed

**Variables**

There should be a common nomenclature for variables used.

See appendix B – to be completed

**Tasks to be included**

Aggregation to ISO-weeks

The underlying time unit is ISO-weeks at week of death. As input has Date-of-Death (DoD), number of deaths per ISO-week must be calculated.

Only full week must be included i.e. with data from all days in the week. Of importance at beginning and end of data.

Delay-adjustment

Delay-adjustment is based on the idea that an equal proportion of total number of deaths is registered every workday, varying with days after death; and has been so over a period of stable registration. This can be cumulated to proportion of deaths expected to have been registered each ISO-week (WoR, Week of Registration) depending of number of workdays in that week.

For example, if in the same week as death occurred 40% of total number of deaths occurring in that week are expected to be registered and 30 deaths has been registered this week, then the delay-adjusted total number will be 30\*(100/40) = 75. If, after 3 weeks 80% of the total amount of deaths in a week are expected to have been registered, and 50 are registered, then the delay-adjusted expected total number of deaths will be 50\*(100/80) = 62.5

In terms that are more general:

proportiondw=0,1,… = ∑ registered at week=dw / ∑ total deaths in dw=0

where proportion is estimated over a period with stable registration and without the last weeks, that should be delay-adjusted

Done in A-MOMO in a binary regression controlling for number of workdays

cnbw = rnb/proportion

where cnb (corrected number of deaths) is estimated for the weeks to be delay-adjusted.

Done in A-MOMO in a Poisson regression using predicted proportion from above

However, we will like to include the uncertainty (variance) in uncertainty when estimating the baseline (see Baseline estimation) (note: this will be 0 for week without delay-adjustment). This imply, that we need the variance of cnb

Will it be possible to make one regression?

cnbw = rnb / ( ∑ registered at week=dw / ∑ total deaths in dw=0 )

Further, we should also consider if it is possible to find number of weeks that need delay-adjustment automatically – I believe so.

Will it be possible automatically to determine needed number of weeks for delay adjustment?

Baseline estimation

Estimation of baseline should be limited to only include the preceding 5 years/seasons and minimum preceding 3. The present season (winter or summer) excluded

Baseline is estimated using a Poisson regression with overdispersion, trend and seasonality:

cnb = α0 + α1 WoDi + α2 sin((2π/ (365.25/7)) WoDi) + α3 cos((2π/ (365.25/7)) WoDi)

and screwness in variance of excess number of deaths (excess = cnb – pnb; pnb = predicted number of deaths) is compensated using a 2/3-power (γ) transformation:

Var(cnbγ- pnbγ) ≈ (γ pnbγ-1)2 (Var(cnb) + Var(pnb))

= (γ pnbγ-1)2 Var(cnb - pnb)

Including the variance from delay-adjustment Var(d), Var(cnb) => Var(cnb) + Var(d), I believe the 2/3-power transformed variance should be:

Var(cnbγ- pnbγ) ≈ (γ pnbγ-1)2 (Var(cnb) + Var(d) + Var(pnb))

= (γ pnbγ-1)2 (Var(cnb - pnb) + Var(d))

= (γ pnbγ-1)2 (Var(cnb - pnb) + (γ pnbγ-1)2 Var(d))

Model evaluation

Residuals over time to evaluate trend

Periodograms to evaluate seasonality – cnb and residuals

Further, the following may be considered. On the other hand, we cannot do anything about it. They are included in Stata A-MOMO

Heteroscedasticity

Residual autocorrelation

Outputs

- tables, with weekly excess and cumulated excess number of deaths by ISO week and season

- graph, with weekly cumulated excess number of deaths by ISO week and season

- graphs, with observed, baseline and 2,4,6,… z-score lines

- graphs, with z-scores

- file to be send to EuroMOMO

- optionally, model evaluation graphs

**APPENDIX A**

Parameters

|  |  |
| --- | --- |
| Country |  |
| Regions/counties | If data not cover the whole country, then region(s) or counties covered |
| Institution |  |
| Working directory | Where data is located |
| Date of Aggregation | The date the local MOMO is run |
| Winter season | Start ISO-week  End ISO-week |
| Input file name | Dates must be ISO-standard (YYYY-MM-DD) |
| BankHoliday file name |  |
| Start of regular deaths registration | Date from which pattern I registration are regular and can be used in the delay-adjustment |
| Delay-adjustment | The number of weeks to be delay-adjusted |
| Previous seasons used in the baseline estimation | max 5, min 3 |
|  |  |
|  |  |
| Groups | The EuroMOMO age groups 0-4, 5-14, 15-64, 65+ must be obligatory, while it also should be possible to make other groups for local use |
|  |  |
| Exception periods | Either a series of start-end ISO-weeks for periods to be excluded in the baseline estimation  Or spring-autumn to be repeated every year |
|  |  |
|  |  |

**APPENDIX B**

Variables

|  |  |
| --- | --- |
| DoD | Date of Deaths |
| DoR | Date of Registration |
| age | Age at death in years |
| DoA | Date of Aggregation – the date of the data. When last updated.  But, also a possibility to run the model back in time |
| agegrp | Age groups: 0-4, 5-14, 15-64, 65+ |
| YoRi | Year of Registration – ISO-week |
| WoRi | Week of Registration – ISO-week |
| YWoRi | Year and Week of Registration – ISO-week |
| YoDi | Year of Deaths – ISO-week |
| WoDi | Week of Deaths – ISO-week |
| YWoDi | Year and Week of Deaths – ISO-week |
| nb | Known number of deaths in that week |
| rnb | Registered number of death in that week |
| cnb | Corrected (delay-adjusted) number of deaths |
| V*x* | Variance of *x* |
| logV*x* | Variance of *x* on the log-scale |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |