Longevitas client webinar

Robust mortality forecasts

Stephen J. Richards

14th February 2023

Overview

LONGEVITAS

1. Motivation
2. Limits of judgement
3. Robust univariate forecasts
4. Robust multivariate forecasts
5. Conclusions

About this presentation

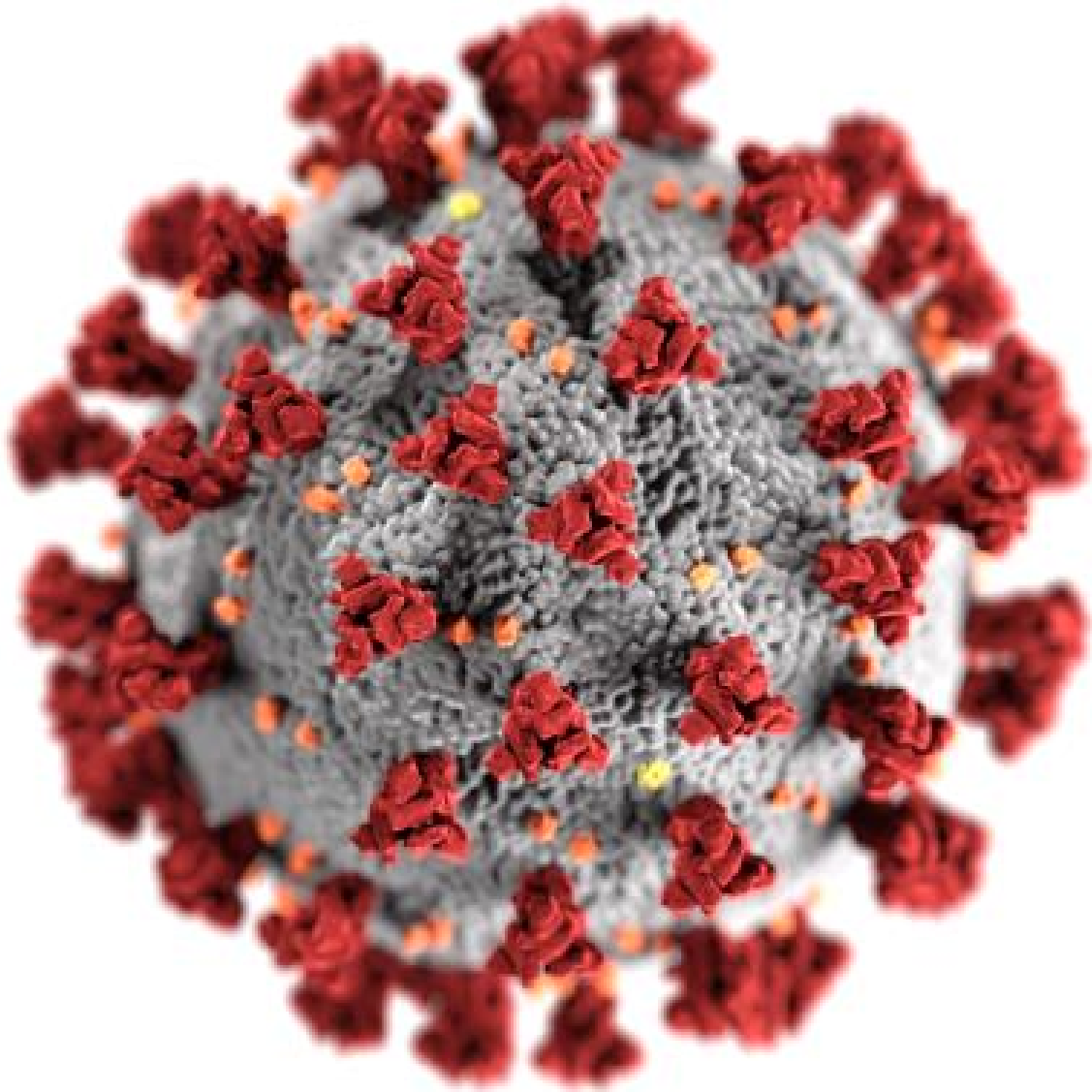
LONGEVITAS

This presentation is about how to use features in the Projections Toolkit software.

For this and other materials on robust forecasting, see

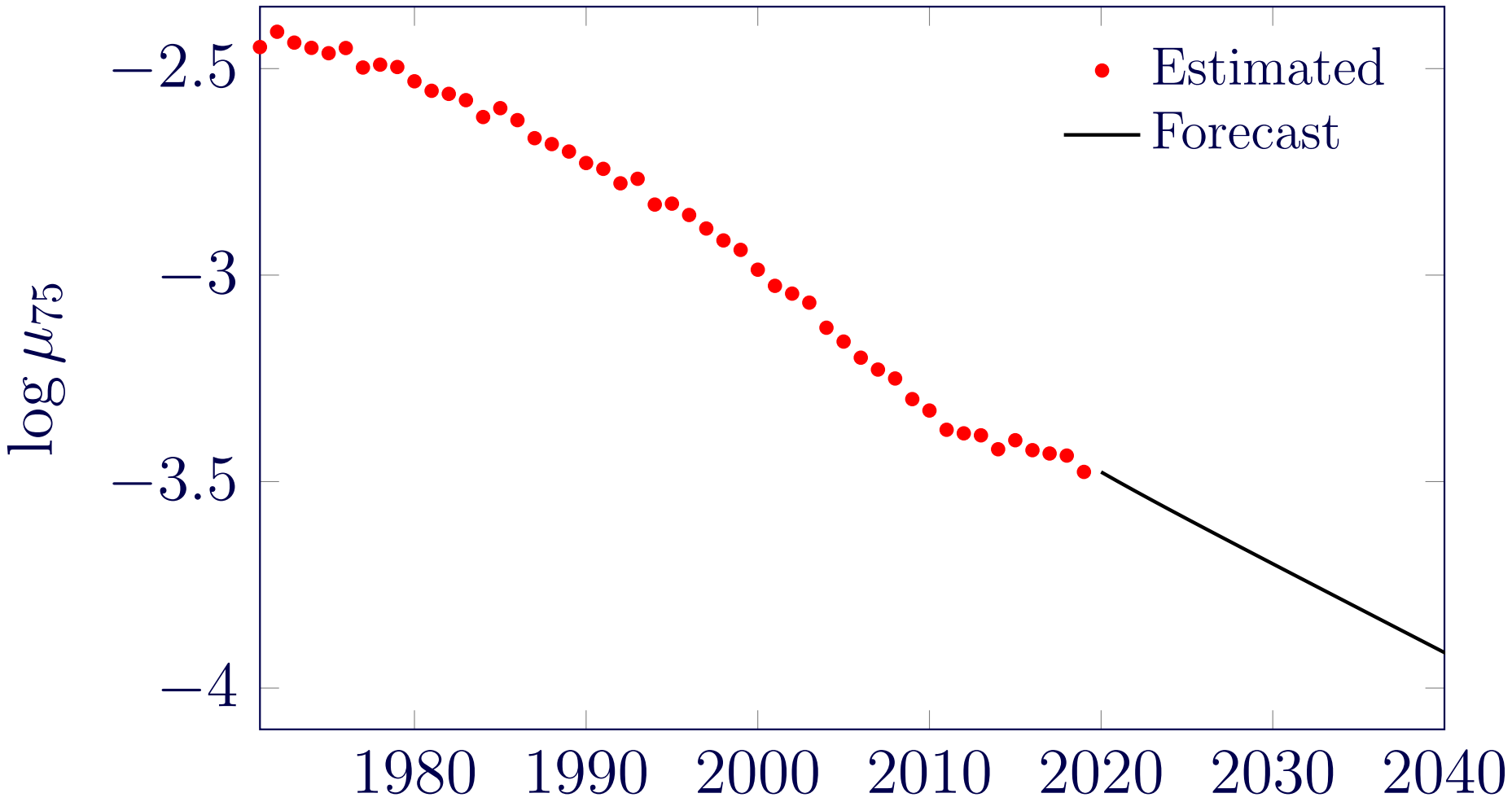
- www.longevitas.co.uk/robust-forecasting

1 Motivation



Using data to 2019

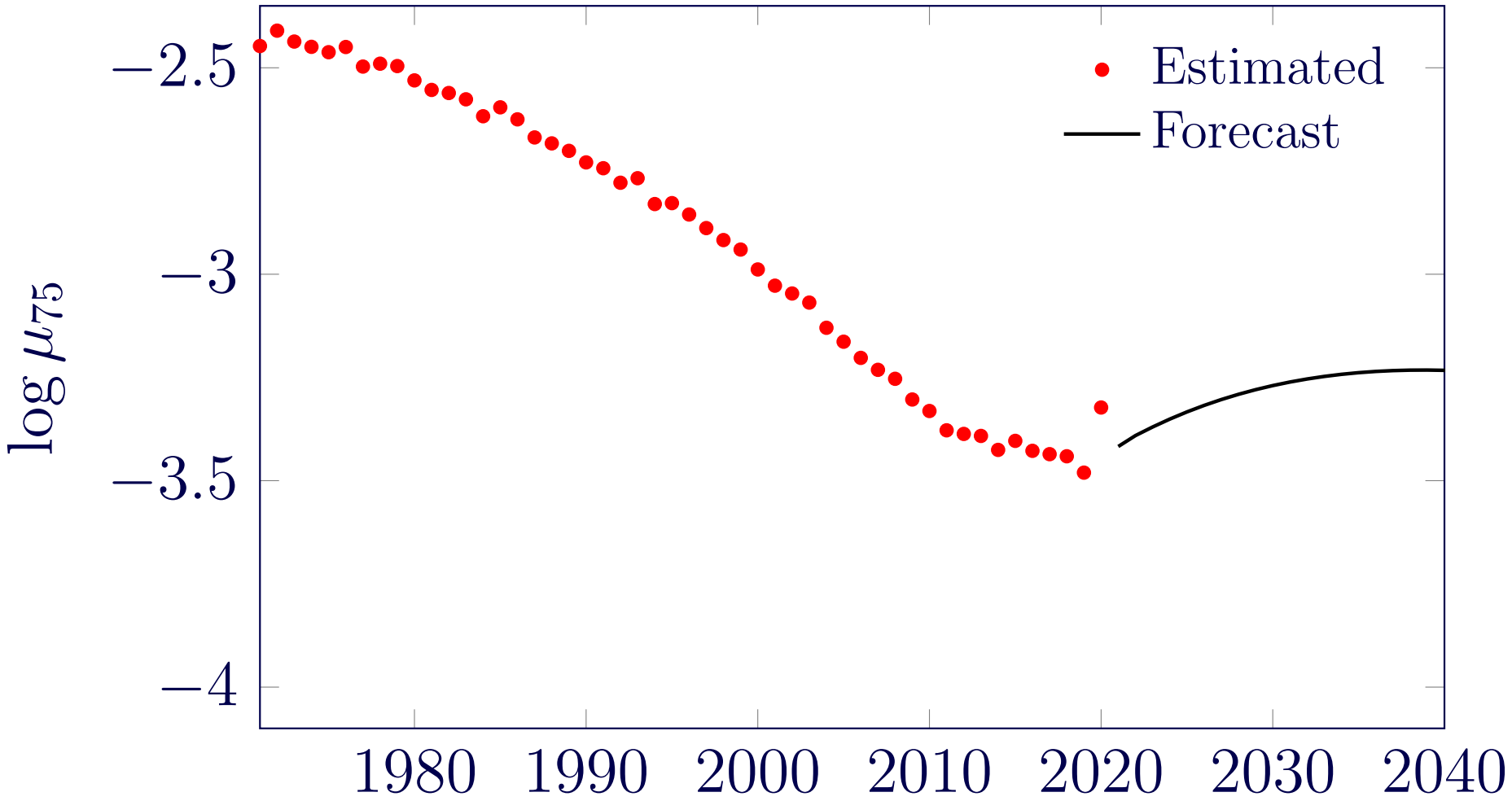
ARIMA forecast of time index in Lee-Carter model:



Source: Data for males in England & Wales, ages 50–105, 1971–2019.

Using data to 2020

ARIMA forecast of time index in Lee-Carter model:



Source: Data for males in England & Wales, ages 50–105, 1971–2020.

1 Problem

- Covid-19 mortality breaks forecasting models.
- How can we robustify actuarial forecasts?

1 Ignore covid-affected years?

- Easy when covid years are at end of data period.
- ✗ Tricky when covid years are in the middle of data period.
- ✗ Subjective.

2 Limits of judgement

The limits of visual inspection

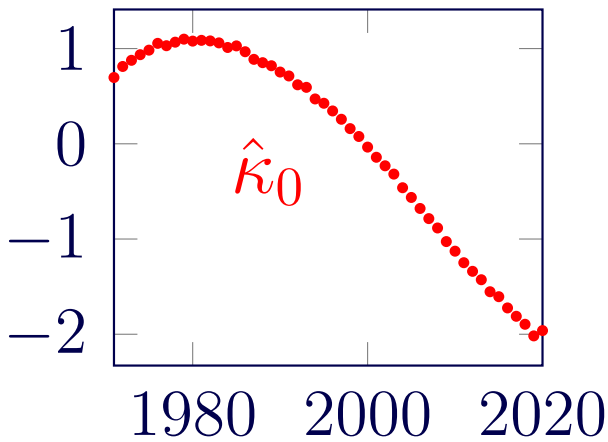
LONGEVITAS

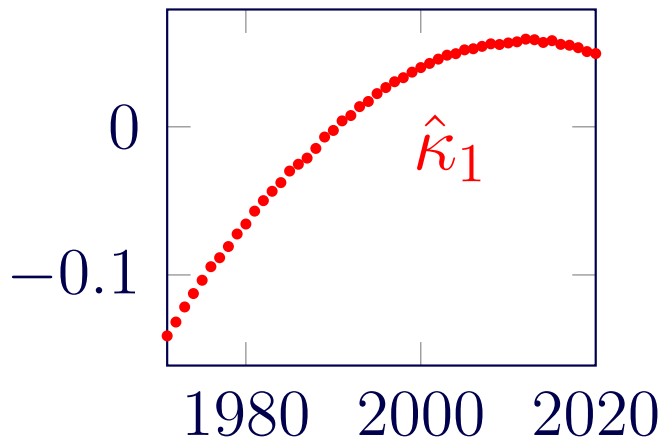
> "good data deserve inspection from a variety of perspectives" Lange [1998]

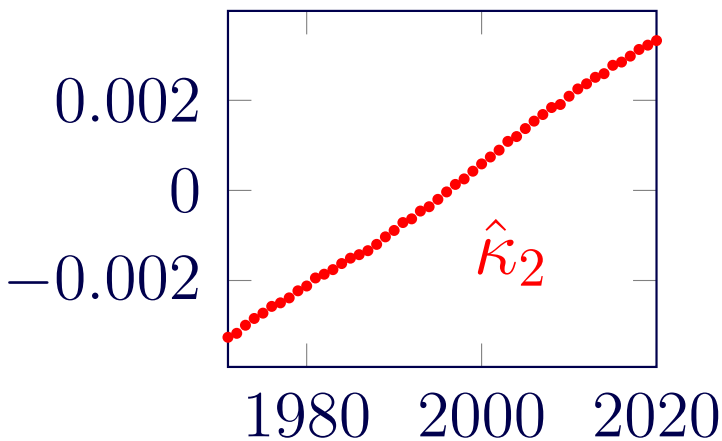
The limits of visual inspection

LONGEVITAS

M9 model [Cairns et al., 2015] fitted to data for males in England & Wales, ages 50–105, years 1971–2020.







Where's the 2020 outlier?

Spotting outliers using plotting

LONGEVITAS

- First differences in CBD time indices are supposed to be i.i.d. multivariate normal.
- Plots are useful for univariate and bivariate random walks.
- Trouble starts with trivariate random walks (M7 and M9), where viewing angle is important...

Spotting outliers using plotting

LONGEVITAS

Depending on the type of plot, 2020 is either:

- An obvious outlier, or
- Not an outlier at all.

2 Requirements & solutions

LONGEVITAS

- Need an objective approach to outliers.

$\Rightarrow$ Identify outliers with statistical tests.

- Need to remove bias and distortion.

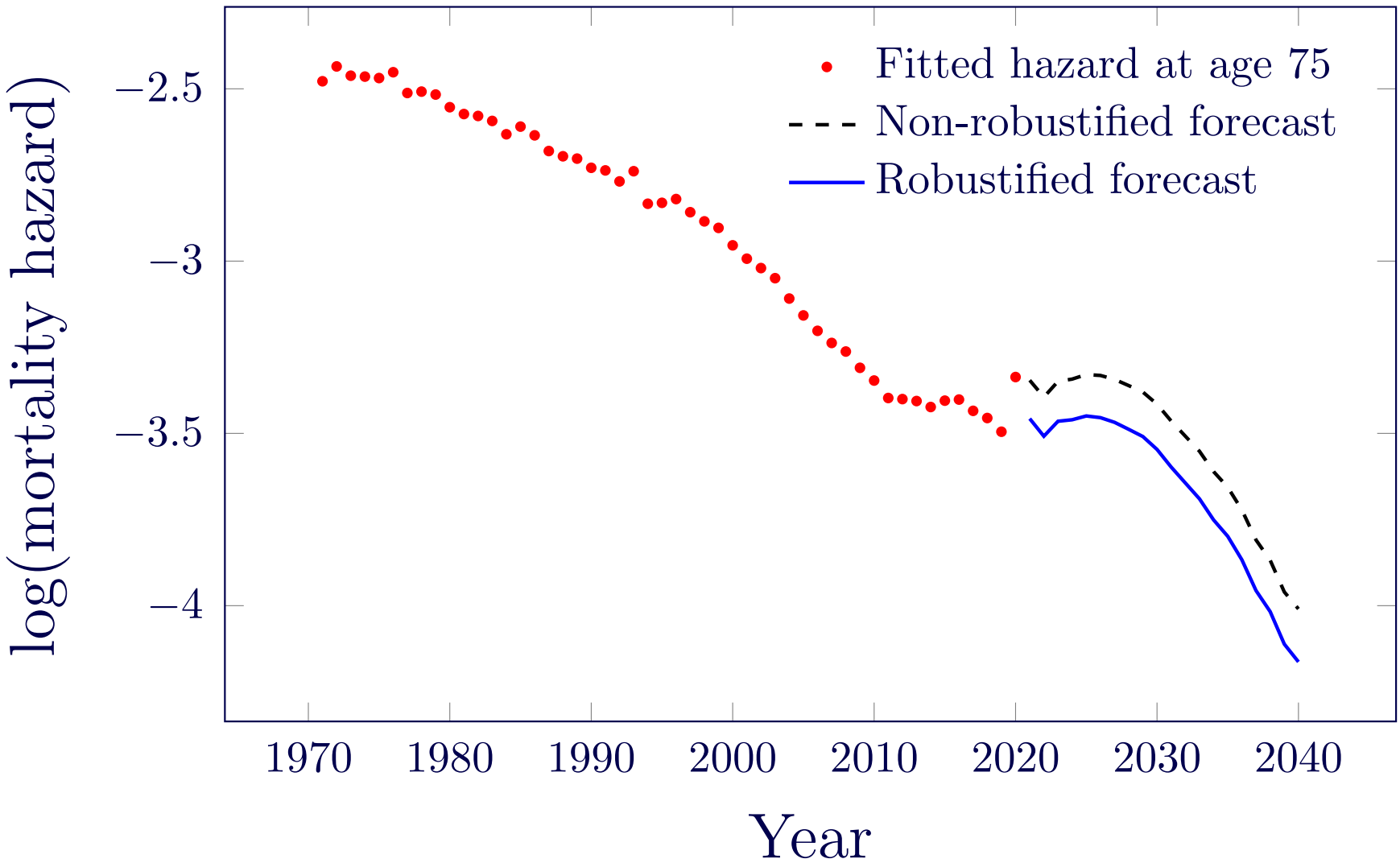
$\Rightarrow$ Co-estimate outlier effects with other parameters.

2 Why estimate outlier effects?

LONGEVITAS

1. Co-estimation of outlier effects removes bias in parameters in forecast model.
2. If the last year is affected, estimation of outlier effects allows calculation of a more sensible starting point for the forecast.

2 Robust starting point



Source: M9 model fitted to data for males in England & Wales.

3 Robust univariate forecasts

- A univariate model has multiple parameter vectors, but only one represents a time index to forecast.
- Example from Lee and Carter [1992]:

$$\log \mu_{x,y} = \alpha_x + \beta_x \kappa_y$$

- $\hat{\alpha}_x$ and $\hat{\beta}_x$ are held constant in the forecast.
- An ARIMA model is fitted to the $\hat{\kappa}_y$ time index to forecast the trend.

3 Outlier definition

LONGEVITAS

Outlier

An observation that is further from the one-year-ahead forecast than is consistent with the noise variance.

To robustify an ARIMA model, Chen and Liu [1993]:

1. Proposed test statistics to identify outliers.
2. Proposed further test statistics to *classify* outliers.

3 Four outlier types

LONGEVITAS

IO	Innovation outlier
AO	Additive outlier
TC	Temporary change
LS	Level shift

3 Innovation outlier

LONGEVITAS

A modest outlier that is nevertheless integrated into the process.

Example

A year with heavy winter mortality due to influenza, possibly with lighter mortality the following year.

Handling: leave alone.

3 Additive outlier

A more extreme outlier that is not integrated into the process.

Example

War or pandemic in a single year.

Handling: co-estimate the outlier effect to remove bias in other parameters.

3 Temporary change

Two or more consecutive outliers that are not integrated into the process.

Example

War or pandemic spread over more than one year.

Handling: co-estimate the outlier effects to remove bias in other parameters.

3 Level shift

Permanent change in level of process.

Example

After German reunification in 1990, old-age mortality in East converged rapidly on levels in the West [Grigoriev et al., 2021].

Handling: review model or data period.

3 Chen and Liu [1993]

LONGEVITAS

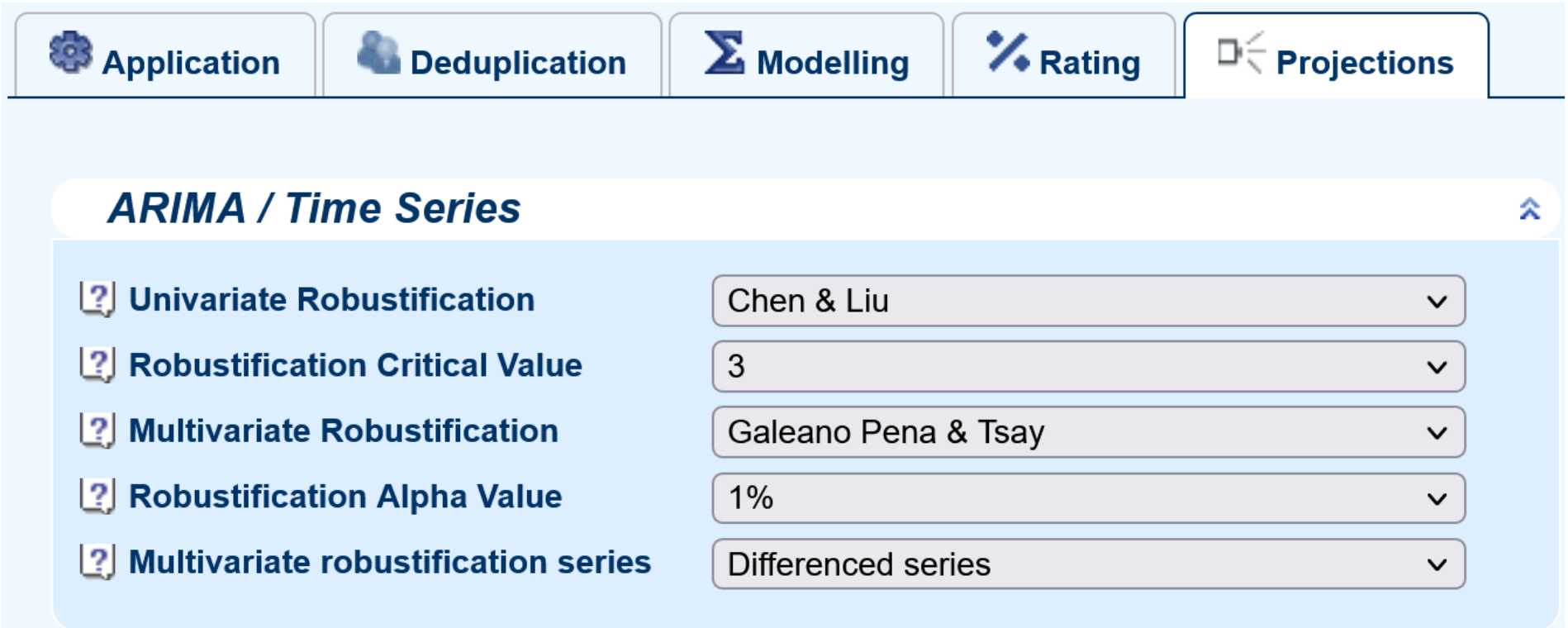
Note

An outlier can be detected anywhere, but its type cannot be determined if it occurs at the end of the series.

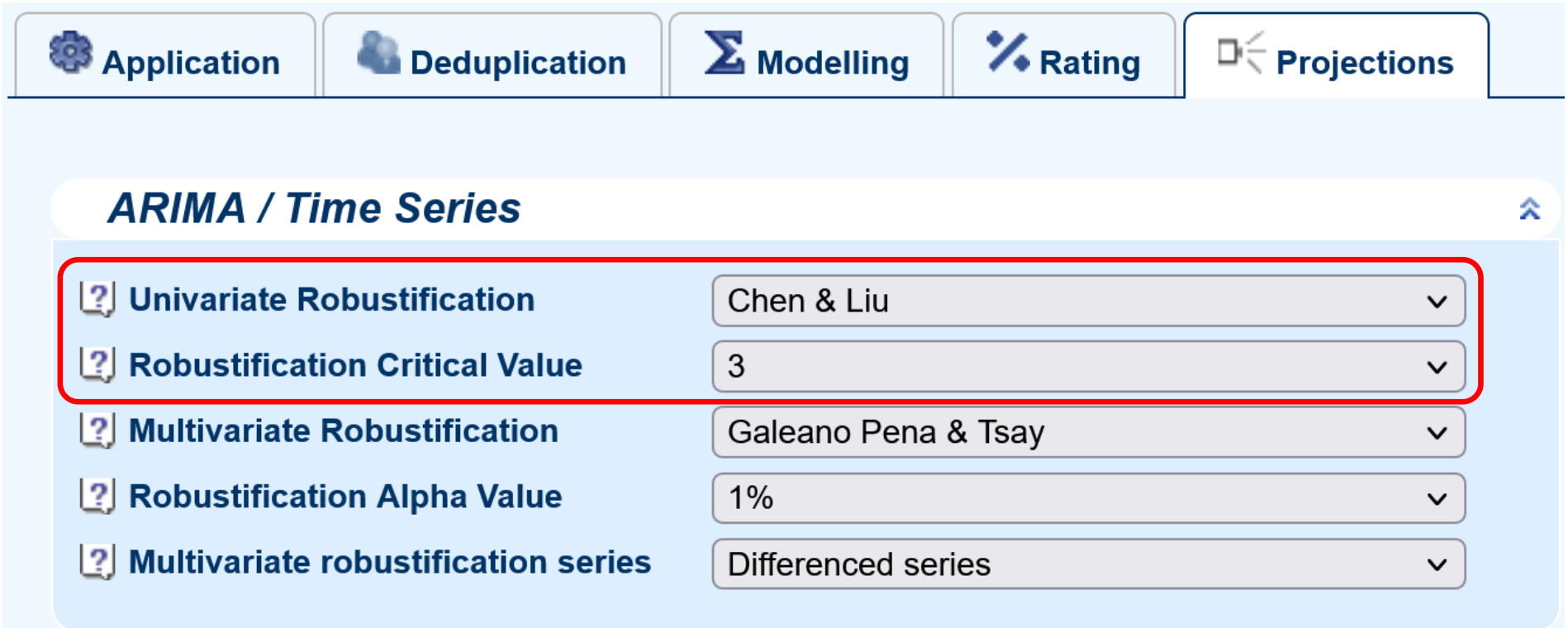
3 Configuration options

LONGEVITAS

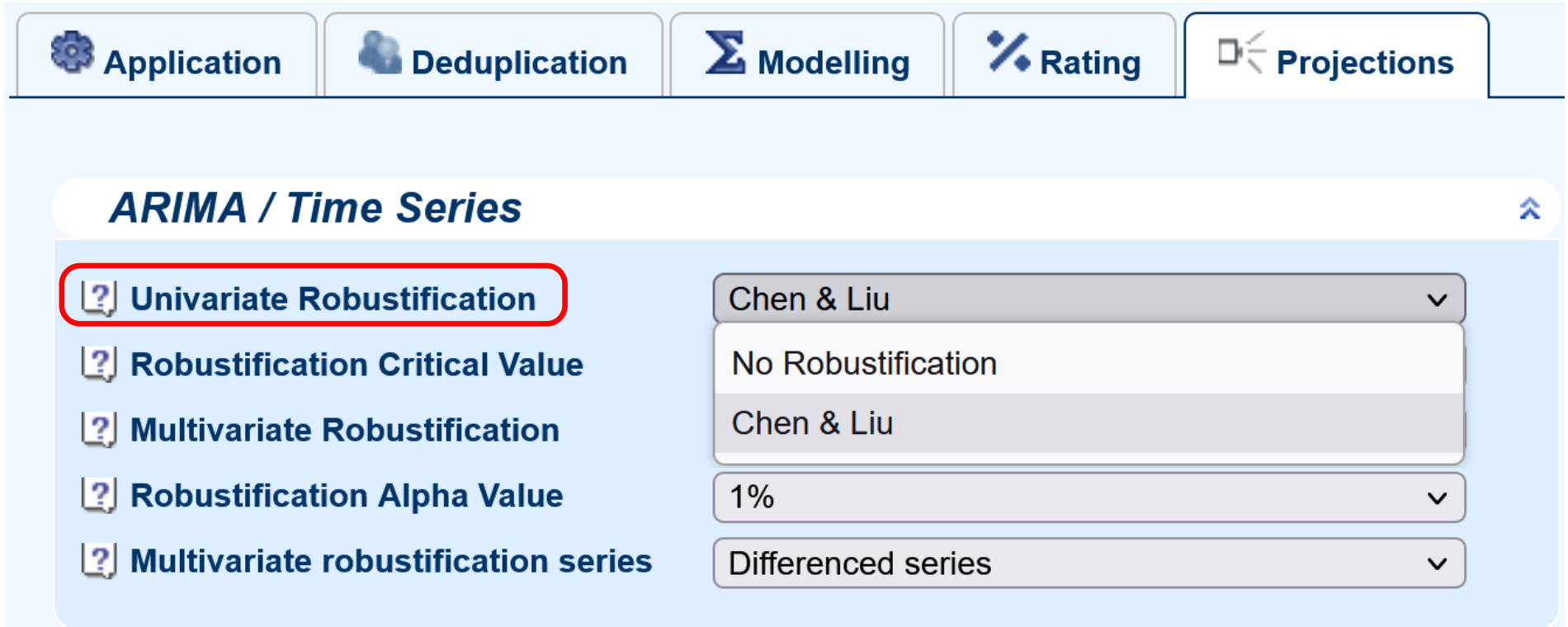
Five controls for robustification:



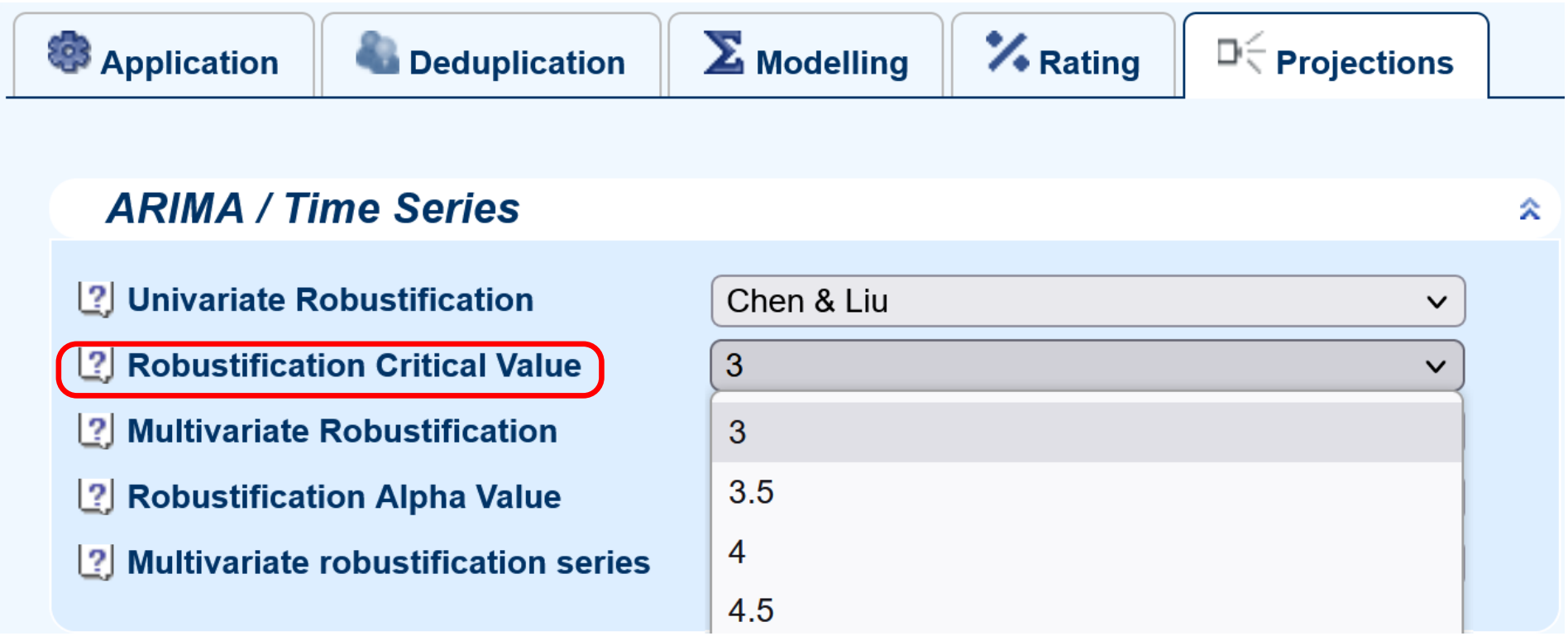
Two controls for univariate robustification:



Switch on Chen and Liu [1993] robustification:



Select critical value for univariate outlier detection:



3 Sensitivity to critical value

Females, E&W, Lee-Carter model, `cval=3`

Summary of estimated outlier effects in undifferenced series with critical value 3.

Year	Outlier type	Estimated outlier effect	t-value
1973	AO	-0.01757	NaN
1976	AO	0.014986	NaN
1980	AO	-0.001498	NaN
1984	AO	-0.017278	NaN
1991	AO	0.029086	NaN
1994	AO	-0.015967	NaN
1996	AO	-0.010621	NaN
1999	AO	-0.000274	NaN
2003	AO	0.025771	130.211535
2007	AO	-0.002924	NaN
2011	AO	-0.033055	NaN
2014	AO	-0.040815	NaN
2018	AO	0.016634	NaN
2020	AO	0.073339	32.687304

Too many outliers

3 Sensitivity to critical value

LONGEVITAS

Females, E&W, Lee-Carter model, `cval=3.5`

Projections

Projections are carried out by means of an ARIMA(0,1,3) time series applied to κ_y. Details of the ARIMA process are shown below. The variance of the noise process, σ^2, is 0.000129.

Summary of ARIMA parameters for κ_y

Parameter	Estimate	Standard error
ma1	-0.842428	0.146775
ma2	0.307704	0.182541
ma3	0.511356	0.133761
mean	-0.009136	0.001568

The ARIMA process is fitted using a Kalman filter *(Harvey, 1996)* via R's *arima()* function. The details of this filter are given in the model PDF report.

Robust estimation of the model for κ_y is carried out using the approach of Chen & Liu (1993), which simultaneously estimates the model parameters alongside outliers.

Summary of estimated outlier effects in undifferenced series with critical value 3.5.

Year	Outlier type	Estimated outlier effect	t-value
2020	AO	0.059706	4.960545

← Sensible number of outliers

4 Robust multivariate forecasts

4 Multivariate forecasting

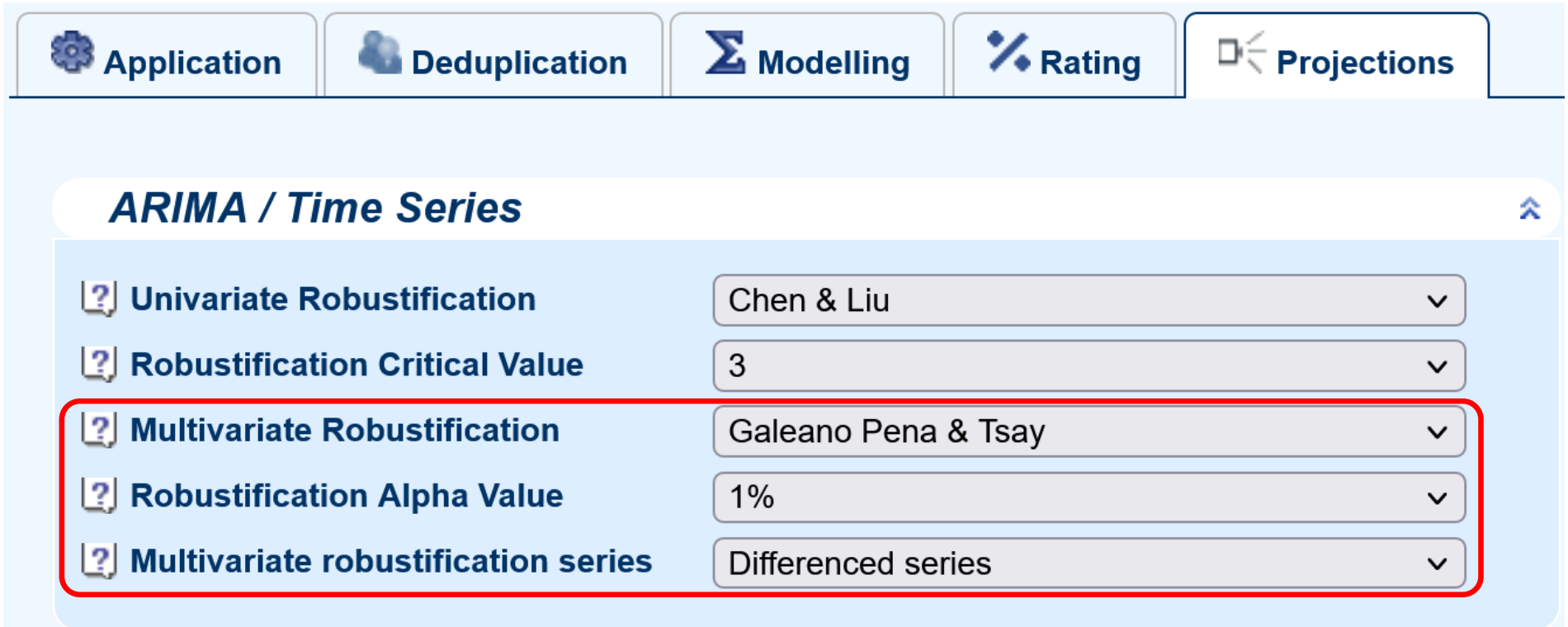
LONGEVITAS

- A multivariate model has two or more time indices.
- Example from Cairns et al. [2006]:

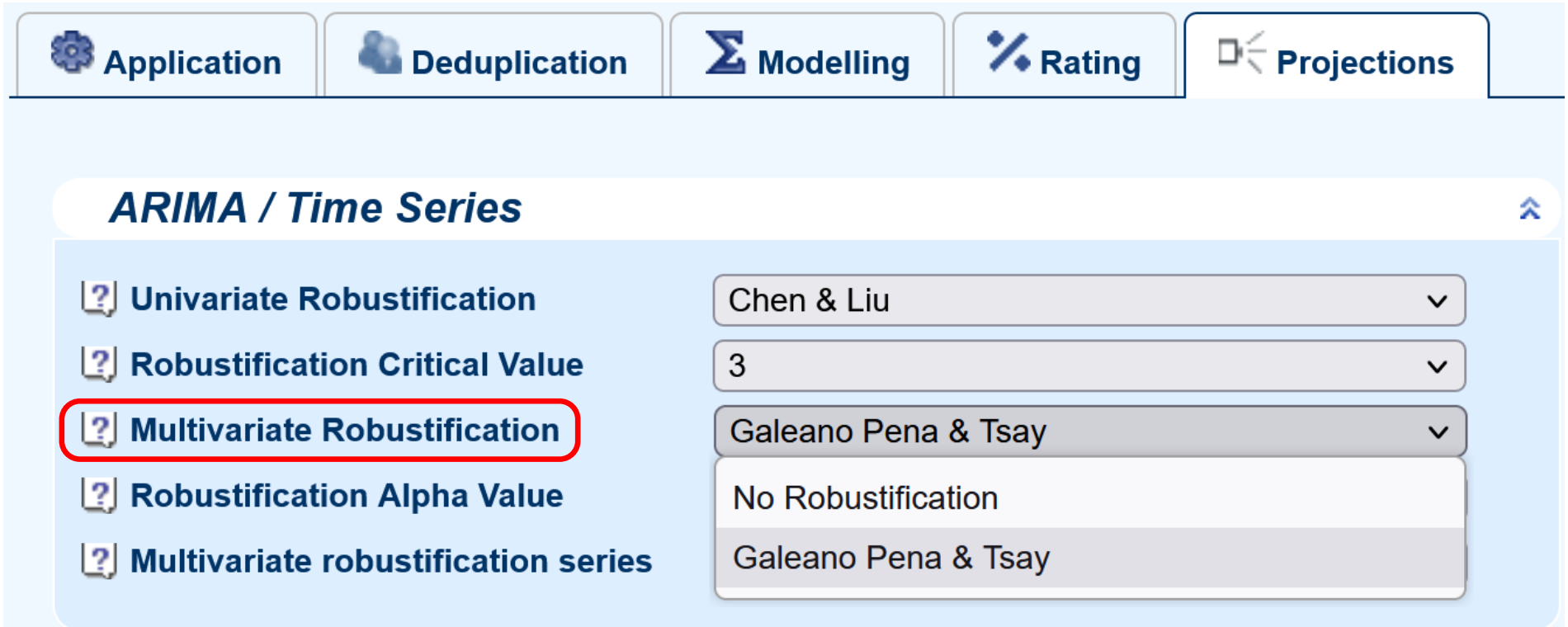
$$\log \mu_{x,y} = \kappa_{0,y} + (x - \bar{x})\kappa_{1,y}$$

- $\hat{\kappa}_{0,y}$ and $\hat{\kappa}_{1,y}$ are forecast jointly as a bivariate random walk with drift.

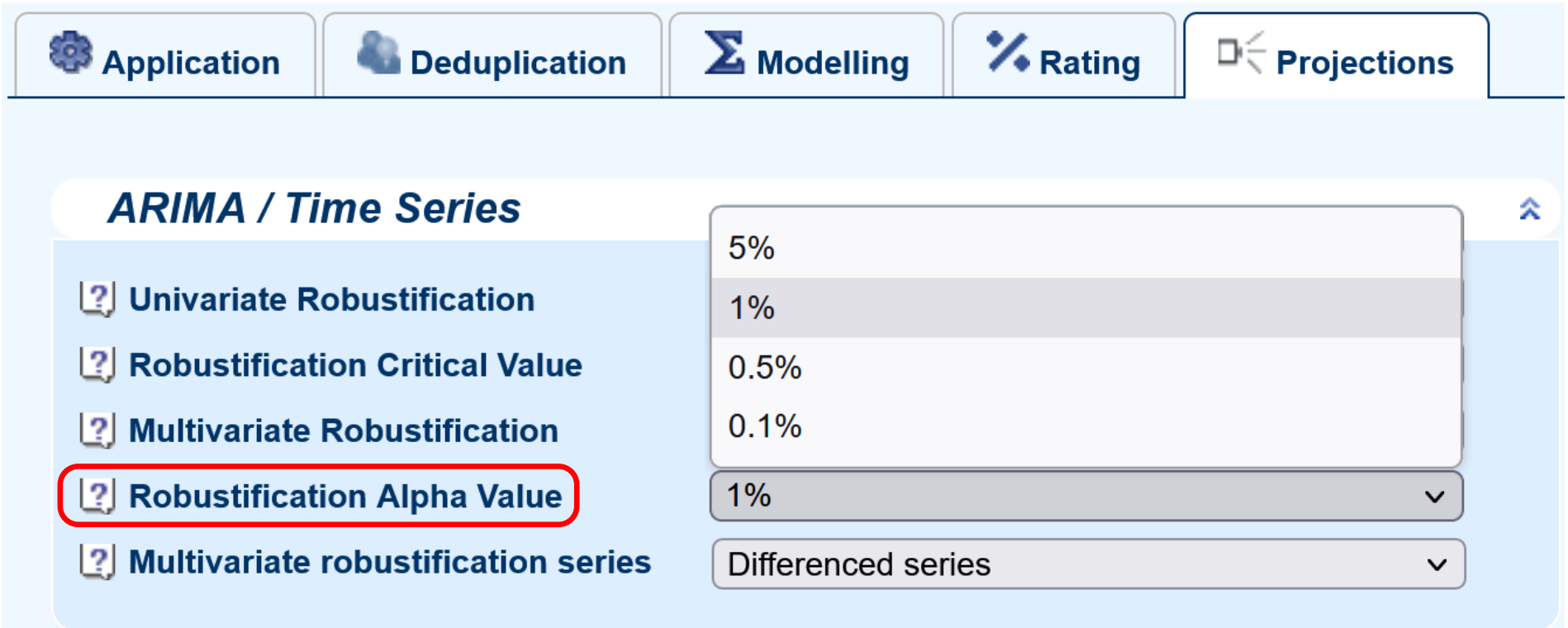
Three controls for multivariate robustification:



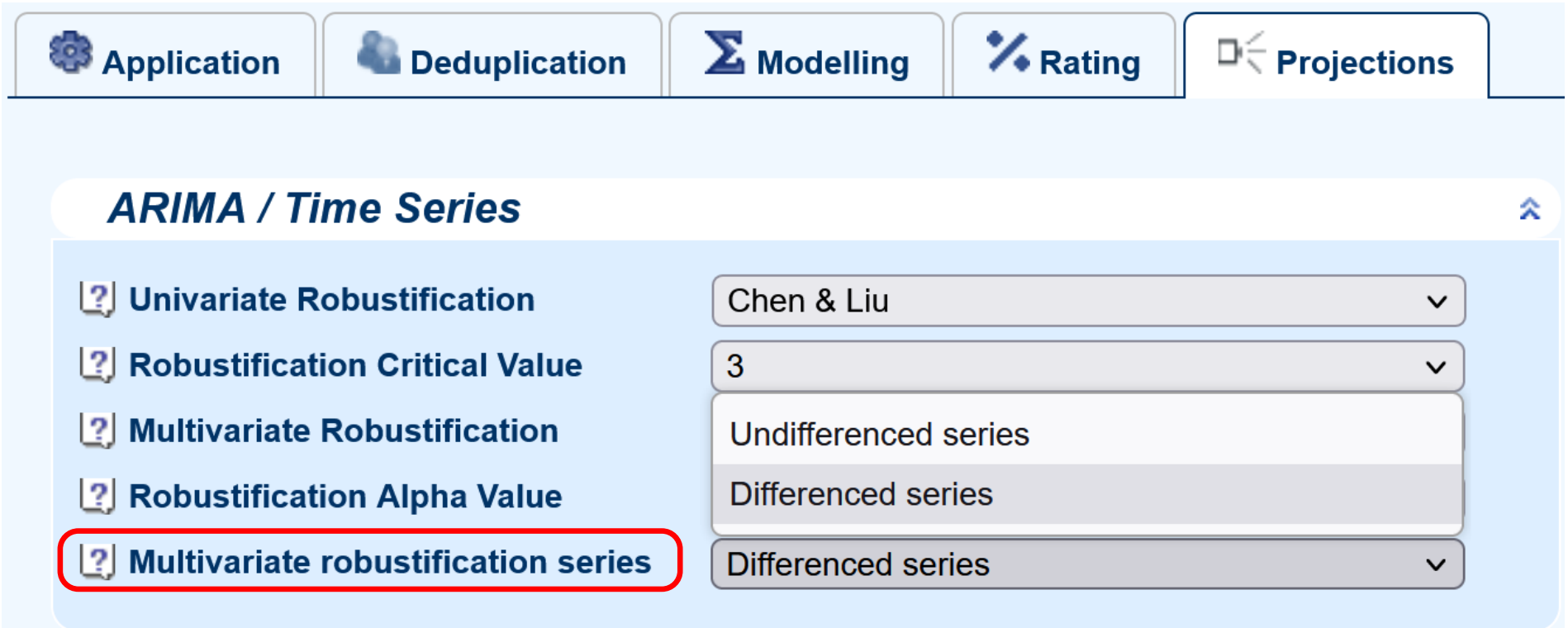
Select Galeano et al. [2006] robustification:



Select the alpha value for outlier detection:



Choose to robustify differenced or undifferenced series:



4 To diff or not to diff?

LONGEVITAS

Projecting κ_0, κ_1 and κ_2

Projections are carried out by means of a trivariate random walk for κ_0, κ_1 and κ_2 with drift parameters given below.

Parameter	Estimate	Standard error	Noise error
κ_0	-0.05727	0.008902	0.062315
κ_1	0.00395	0.00049	0.003431
κ_2	0.000131	0.000007	0.000048

The standard deviations of the noise process ("innovations"), and the correlation matrix between the drift terms are shown in the model PDF report.

Robust estimation of the model for κ_y is carried out using the approach of Galeano, Peña & Tsay (2006), which simultaneously estimates the forecast model parameters alongside identification of outliers and their effects. Robustification is applied to the undifferenced series using a p-value of 1% as the threshold. Below is a table of the estimated outlier effects:

Year	Estimate of outlier effect
2020	(0.148798, -0.002931, 0.000184)

4 To diff or not to diff?

LONGEVITAS

Projecting κ_0, κ_1 and κ_2

Projections are carried out by means of a trivariate random walk for κ_0, κ_1 and κ_2 with drift parameters given below.

Parameter	Estimate	Standard error	Noise error
κ_0	-0.057557	0.008931	0.062514
κ_1	0.003875	0.000503	0.003523
κ_2	0.000135	0.000005	0.000036

The standard deviations of the noise process ("innovations"), and the correlation matrix between the drift terms are shown in the model PDF report.

Robust estimation of the model for κ_y is carried out using the approach of Galeano, Peña & Tsay (2006), which simultaneously estimates the forecast model parameters alongside identification of outliers and their effects. Robustification is applied to the differenced series using a p-value of 1% as the threshold. Below is a table of the estimated outlier effects:

Year	Estimate of outlier effect
2020	(0.162825, 0.000751, -0.000011)

4 Other multivariate options

LONGEVITAS

Other approaches to identifying outliers in multivariate data, e.g. Hadi [1994].

5 Conclusions

5 Conclusions — I

- Covid-19 distorts mortality forecasting models.
- Ignoring affected years only a temporary solution.
- Need objective statistical tests for outliers.

5 Conclusions — II

LONGEVITAS

- Co-estimation of outliers and parameters reduces bias in latter.
- Also permits calculation of better starting point for forecasts.

5 Conclusions — III

LONGEVITAS

Co-estimate outlier effects and model parameters:

Univariate forecasting

- Lee-Carter, APC, APCI and Age-Period.
- Use approach of Chen and Liu [1993].

Multivariate forecasting

- Cairns-Blake-Dowd family: M5, M6, M7 and M9.
- Use approach of Galeano et al. [2006].

References I

A. J. G. Cairns, D. Blake, and K. Dowd. A two-factor model for stochastic mortality with parameter uncertainty: theory and calibration. *Journal of Risk and Insurance*, 73:687–718, 2006. doi: 10.1111/j.1539-6975.2006.00195.x.

A.J.G. Cairns, D. Blake, K. Dowd, and A. Kessler. Phantoms never die: Living with unreliable mortality data. *Journal of the Royal Statistical Society, Series A*, 2015. doi: 10.1111/rssa.12159.

References II

C. Chen and L-M. Liu. Joint estimation of model parameters and outlier effects in time series. *Journal of the American Statistical Association*, 88(421): 284–297, 1993.

P. Galeano, D. Peña, and R. S. Tsay. Outlier detection in multivariate time series by projection pursuit. *Journal of the American Statistical Association*, 101 (474):654–669, 2006. doi: 10.1198/016214500001131.

References III

P. Grigoriev, M. Pechholdová, M. Mühlichen, R. D. Scholz, and S. Klüsener. 30 Jahre Deutsche Einheit: Errungenschaften und verbliebene Unterschiede in der Mortalitätsentwicklung nach Alter und Todesursachen. *Bundesgesundheitsblatt*, pages 481–490, 2021. doi: 10.1007/s00103-021-03299-9.

A. S. Hadi. A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society, Series B (Methodological)*, 56(2):393–396, 1994. ISSN 00359246. URL `http://www.jstor.org/stable/2345910`.

References IV

K. Lange. *Numerical Analysis for Statisticians*. Springer, 1998. ISBN 0-387-94979-8.

R. D. Lee and L. Carter. Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87:659–671, 1992. ISSN 01621459. URL `http://www.jstor.org/stable/2290201`.

Coronavirus graphic  from CDC

More on robust forecasting at ▸ www.longevitas.co.uk/robust-forecasting

6 Legal matters

Longevitas is a registered trademark:

- in the UK (No. 2434941),
- in the USA (No. 3707314), and
- in the European Union (No. 5854518).