AVERISERA FORECASTING TOOL MANUAL

VERSION 1.0

INTRODUCTION

Averisera Forecasting Tool is an add-in for Microsoft Excel® which provides functions for interpolating and extrapolating time-dependent distributions of categorical variables based on repeated cross-sectional data. It implements two methods:

- **Cross-sectional Markov model** (CSM) developed by Averisera Ltd, described in our paper *Cross-sectional Markov model for trend analysis of observed discrete distributions of population characteristics*, https://arxiv.org/abs/1510.06787
- **Multinomial logistic regression** (MLR), see e.g.: https://en.wikipedia.org/wiki/Multinomial_logistic_regression

INSTALLATION IN MICROSOFT WINDOWS®

1. Create directory for the add-in, e.g. `C:\Averisera`
2. Add the full path of the `C:\Averisera` to `PATH` environment variable
   The fastest way to do this is typing "path" in Windows' Start Menu search box, selecting "Edit environment variables for your user account", and editing variable "Path" by adding `C:\Averisera` (preceded by a semicolon) to its value.
3. Extract all files from the downloaded archive `ForecastingTool.zip` to `C:\Averisera`
4. If you want to use the 32-bit version of the add-in, rename the file `libnlopt-0_32bit.dll` to `libnlopt-0.dll` and `boost_date_time-vc140-mt-gd-1_59_32bit.dll` to `boost_date_time-vc140-mt-gd-1_59.dll`. You can store the 64-bit DLL files (without the `_32bit` suffix) in another directory in case you decide to switch to the 64-bit add-in version later.
5. Save the licence file `addin.lic` that you have received by email (to the email address provided by you in the purchase process) to `C:\Averisera`
6. Run Excel and install the add-in in the following steps:
   a. Go to `File/Options/Add-ins`
   b. In "Manage" drop-down list choose "Excel Add-ins"
   c. Click "Go…"
   d. Click "Browse…"
   e. Navigate to `C:\Averisera` and select the file `ForecastingTool32.xll` or `ForecastingTool64.xll`, depending on whether you use a 32-bit or 64-bit version of Excel

7. The add-in should load and display "Averisera Forecasting Tool © Averisera Ltd 2017; licensed to *username@domain.com* (licence ID *1*)" in the Excel status bar (where *username@domain.com* is the email address that you provided in the purchase process, and the number after "licence ID" identifies the particular licence you hold).
8. Open the included example file `ForecastingToolExample.xlsx` to learn how to use the new functions.

In the case of any problems with the installation, please contact Averisera Ltd at info@averisera.uk

USING THE EXTRAPOLATION MODELS FOR PROBABILITY DISTRIBUTIONS

INPUT DATA

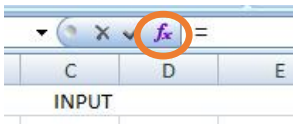
`ForecastingToolExample.xlsx` presents an example of a trend analysis of repeated cross-sectional data divided into three categories (3-dimensional time-dependent distributions). The input data has the following layout:

	A	B	C	D	E	F
1			INPUT			
2						
3	Year		Probabilities		Nbr. surveys	
4	1980	10%	40%	50%	800	
5	1981	12%	39%	49%	810	
6	1982	14%	38%	48%	750	
7	1984	16%	37%	47%	880	
8	1986	18%	36%	46%	910	
9	1987	20%	35%	45%	690	
10	1988	22%	34%	44%	550	
11	1989	24%	33%	43%	670	
12	1991	28%	31%	41%	720	
13	1992	30%	30%	40%	790	

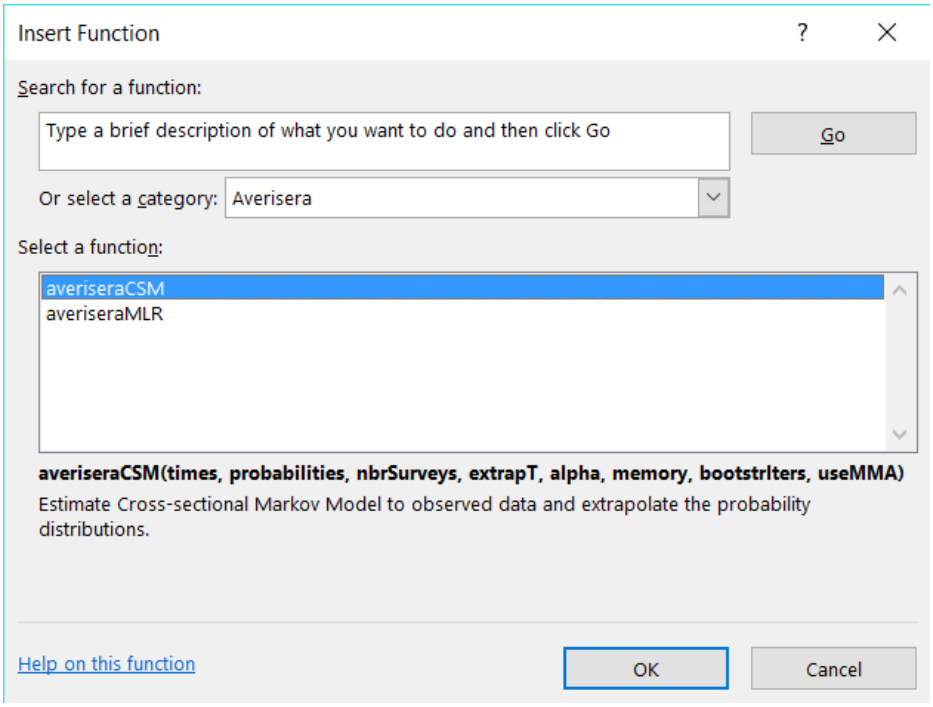
The year, observed probability distribution and number of surveys (optional) are laid out in rows.

CHOOSING THE MODEL

`ForecastingToolExample.xlsx` contains two tabs, CSM and MLR, with calls to respective functions in cell **G2**. To create them we selected the cell **G2**, opened the "Insert function" dialog:



selected the category "Averisera" from the drop-down menu:



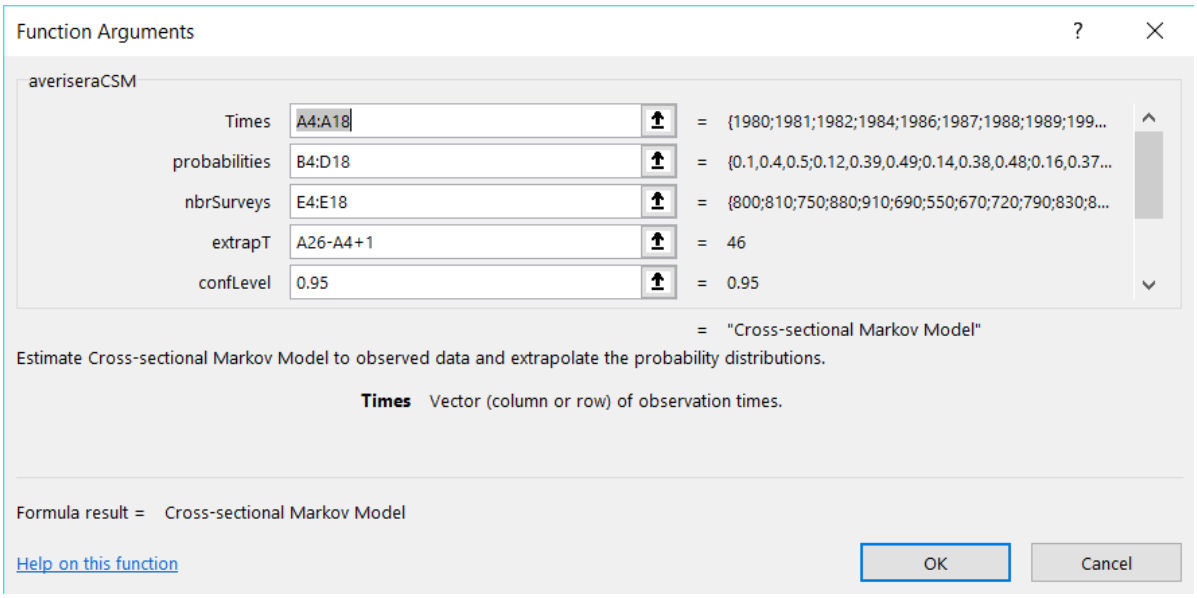
and choose a function corresponding to one of the two methods implemented in the current version of the add-in:

- averiseraCSM – Averisera Cross-sectional Markov model (CSM)
- averiseraMLR – Multinomial Logistic Regression (MLR)

Pressing "OK" opens the dialog "Function Arguments", which asks you to provide arguments and parameters for the selected function, as described in the next section.

FUNCTION PARAMETERS

To display the function parameters used in `ForecastingToolExample.xlsx`, select the cell **G2** in one of the tabs, e.g. CSM, and click the icon “Insert function”:



averiseraCSM

Parameter	Description
Times	Vector (column or row) of observation times
probabilities	Matrix with observed probability distributions (in rows)
nbrSurveys	Vector (column or row) of numbers of surveys (OPTIONAL; defaults to 1 survey in every year)
extrapT	Number of extrapolation periods (OPTIONAL; defaults to 0, i.e. interpolation only)
confLevel	Confidence level, 1 – α, for confidence intervals. Must be larger than 0 and lower than 1 (OPTIONAL; defaults to 0.95)
Memory	Memory length (OPTIONAL; defaults to 0)
bootstrIters	Number of bootstrapping iterations; pass 0 for analytic confidence intervals (OPTIONAL; defaults to 0)
useMMA	Use the Method of Moving Asymptotes[1] optimisation algorithm to obtain an initial guess for transition matrix (if True), or rely on the model's guessing algorithm (if False). OPTIONAL; defaults to True

averiseraMLR

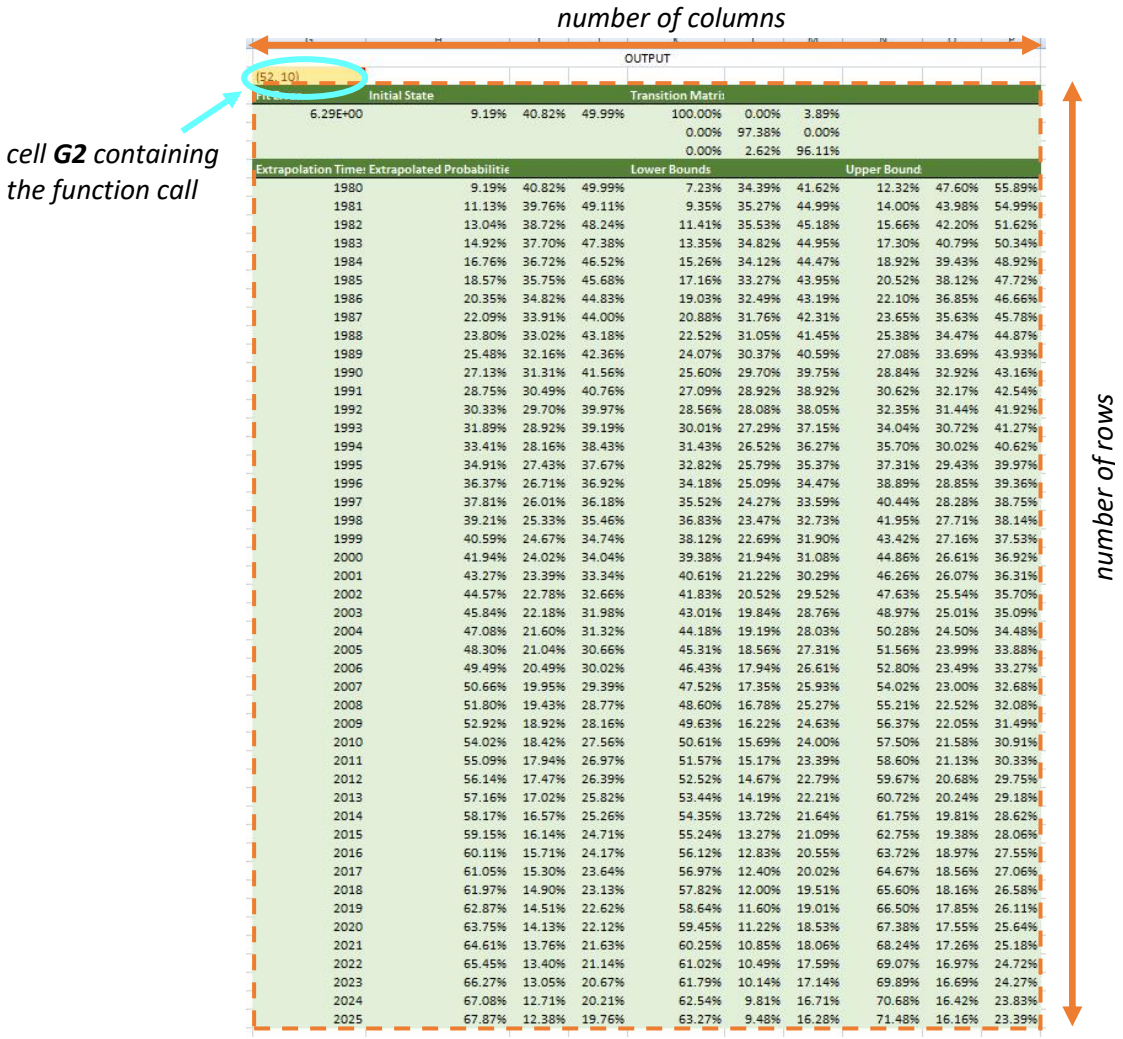
Parameter	Description
Times	Vector (column or row) of observation times
probabilities	Matrix with observed probability distributions (in rows)
nbrSurveys	Vector (column or row) of numbers of surveys (OPTIONAL; defaults to 1 survey in every year)
extrapT	Vector (column or row) of extrapolation times (OPTIONAL; defaults to 0, i.e. interpolation only)
confLevel	Confidence level, 1 – α, for confidence intervals. Must be larger than 0 and lower than 1 (OPTIONAL; defaults to 0.95)

Clicking "OK" executes the function with the provided data and parameters.

DISPLAYING RESULTS

Both functions averiseraCSM and averiseraMLR return the results in the form of an array containing years, estimated probabilities and their confidence intervals, as well as the fitted parameters of models they employ. To see these results, you need to select the range of cells in which they will be displayed. This output range has to be selected and calculated according to the following rules, as explained on the following example of `ForecastingToolExample.xlsx` file:

[1] See http://ab-initio.mit.edu/wiki/index.php/NLopt_Algorithms#MMA_.28Method_of_Moving_Asymptotes.29_and_CCSA



Calculate how many rows and columns is required for the output given the observed distribution dimension *D*, memory of the process *M* (applicable to averiseraCSM only) and number of extrapolation periods *T*, according to the following formulas:

for averiseraCSM	for averiseraMLR
○ *number of rows* = $T + D^{M+1} + 3$ ○ *number of columns* = ▪ $3 \cdot D + 1$ for $M = 0$ ▪ $2 \cdot D^{M+1} + 1$ for $M \geq 1$	○ *number of rows* = $T + 2 \cdot (D - 1) + 3$ ○ *number of columns* = ▪ $4 \cdot (D - 1) + 1$ for $D \leq 4$ ▪ $3 \cdot D + 1$ for $D > 4$

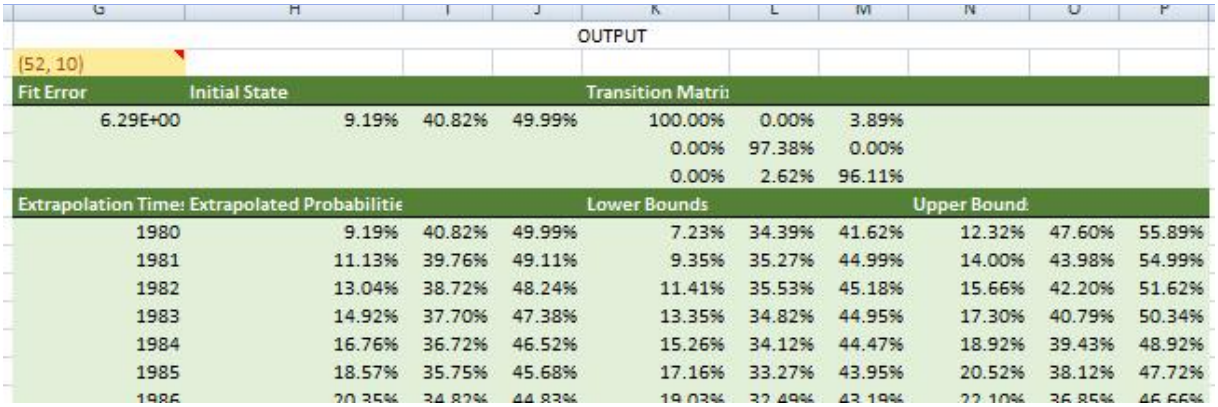
After selecting the required output area click "Insert function" and then press Ctrl + Shift + Enter to execute the function. Once the calculation is finished, the results fill in the output cells. Note that the cell **G2** containing the function call now displays the total number of rows and columns of the returned array in the format (*number of rows*, *number of columns*).

If the results are clipped because they do not fit in the selected output area, you can re-expand it by selecting a larger range, pressing F2 and then Ctrl + Shift + Enter. Warning: doing so will execute the function again (model functions do not cache their results).

RETURNED VALUES

For a detailed explanation of the returned results see our paper *Cross-sectional Markov model for trend analysis of observed discrete distributions of population characteristics*, https://arxiv.org/abs/1510.06787

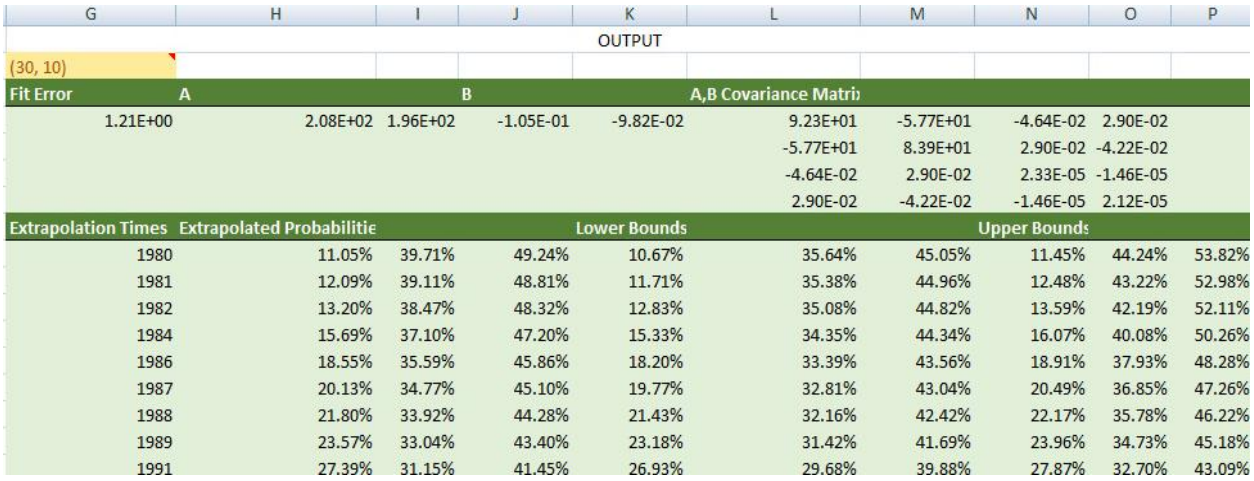
averiseraCSM



G	H	I	J	K	L	M	N	O	P
				OUTPUT					
(52, 10)									
Fit Error	Initial State			Transition Matri					
6.29E+00	9.19%	40.82%	49.99%	100.00%	0.00%	3.89%			
				0.00%	97.38%	0.00%			
				0.00%	2.62%	96.11%			
Extrapolation Time:	Extrapolated Probabilitie			Lower Bounds			Upper Bound:		
1980	9.19%	40.82%	49.99%	7.23%	34.39%	41.62%	12.32%	47.60%	55.89%
1981	11.13%	39.76%	49.11%	9.35%	35.27%	44.99%	14.00%	43.98%	54.99%
1982	13.04%	38.72%	48.24%	11.41%	35.53%	45.18%	15.66%	42.20%	51.62%
1983	14.92%	37.70%	47.38%	13.35%	34.82%	44.95%	17.30%	40.79%	50.34%
1984	16.76%	36.72%	46.52%	15.26%	34.12%	44.47%	18.92%	39.43%	48.92%
1985	18.57%	35.75%	45.68%	17.16%	33.27%	43.95%	20.52%	38.12%	47.72%
1986	20.35%	34.82%	44.83%	19.03%	32.49%	43.19%	22.10%	36.85%	46.66%

Fit Error	Sum of Kullback-Leibler divergences measuring the difference between observed and predicted probability distributions, weighted by the number of surveys in each period
Initial State	Distribution fitted for first observation time
Transition Matrix	Describes the conditional probability of next process value
Extrapolation Times	Times for which the probability distributions were calculated
Extrapolated Probabilities	Probability distributions estimated by the CSM model
Upper / Lower Bounds	Upper and lower confidence intervals (analytical or boostrapped)

averiseraMLR



G	H	I	J	K	L	M	N	O	P
				OUTPUT					
(30, 10)									
Fit Error	A		B		A,B Covariance Matri				
1.21E+00	2.08E+02	1.96E+02	-1.05E-01	-9.82E-02	9.23E+01	-5.77E+01	-4.64E-02	2.90E-02	
					-5.77E+01	8.39E+01	2.90E-02	-4.22E-02	
					-4.64E-02	2.90E-02	2.33E-05	-1.46E-05	
					2.90E-02	-4.22E-02	-1.46E-05	2.12E-05	
Extrapolation Times	Extrapolated Probabilitie			Lower Bounds			Upper Bounds		
1980	11.05%	39.71%	49.24%	10.67%	35.64%	45.05%	11.45%	44.24%	53.82%
1981	12.09%	39.11%	48.81%	11.71%	35.38%	44.96%	12.48%	43.22%	52.98%
1982	13.20%	38.47%	48.32%	12.83%	35.08%	44.82%	13.59%	42.19%	52.11%
1984	15.69%	37.10%	47.20%	15.33%	34.35%	44.34%	16.07%	40.08%	50.26%
1986	18.55%	35.59%	45.86%	18.20%	33.39%	43.56%	18.91%	37.93%	48.28%
1987	20.13%	34.77%	45.10%	19.77%	32.81%	43.04%	20.49%	36.85%	47.26%
1988	21.80%	33.92%	44.28%	21.43%	32.16%	42.42%	22.17%	35.78%	46.22%
1989	23.57%	33.04%	43.40%	23.18%	31.42%	41.69%	23.96%	34.73%	45.18%
1991	27.39%	31.15%	41.45%	26.93%	29.68%	39.88%	27.87%	32.70%	43.09%

Fit Error	Defined in the same way as for the CSM model
A, B	Vectors of MLR parameters, as defined by the MLR equation:

	$P(X_t = k) = \frac{1}{1 + \sum_{l=0}^{D-2} \exp(A_l + B_l t)} \times \begin{cases} 1 & k = 0 \\ \exp(A_{k-1} + B_{k-1} t) & k > 0 \end{cases}$
Covariance Matrix	Covariance matrix of A and B parameters
Extrapolation Times, Extrapolated Probabilities	Defined in the same way as for averiseraCSM function
Lower / Upper Bounds	Upper and lower confidence intervals (averiseraMLR supports **only analytical confidence intervals**)

THIRD-PARTY INTELLECTUAL PROPERTY AND TRADEMARKS

Averisera Forecasting Tool uses the following libraries licensed under the LGPL licence:

1. NLopt http://ab-initio.mit.edu/wiki/index.php/NLopt
2. Sacado https://trilinos.org/packages/sacado/

You can access the text of the LGPL licence under https://www.gnu.org/copyleft/lesser.html

Averisera Forecasting Tool uses the XLW library http://xlw.sourceforge.net/ licensed under the modified BSD licence (http://xlw.sourceforge.net/license.shtml).

Averisera Forecasting Tool uses the f2c library http://www.netlib.org/f2c/ licensed under the AT&T licence (see f2c source).

Excel® and Windows® are registered trademarks of Microsoft Corporation.